Georgia State University
ScholarWorks @ Georgia State University

Mathematics Dissertations | Department of Mathematics and Statistics

12-15-2016

Novel Nonparametric Methods For ROC Curves

Yueheng An

NOVEL NONPARAMETRIC METHODS FOR ROC CURVES

by

YUEHENG AN

Under the Direction of Yichuan Zhao, PhD

ABSTRACT

The receiver operating characteristic (ROC) curve is a widely used graphical method for evaluating the discriminating power of a diagnostic test or a statistical model in various areas such as epidemiology, industrial quality control and material testing, etc. One important quantitative measure summarizing the ROC curve is the area under the ROC curve (AUC). The accuracy of two diagnostic tests with right censored data can be compared using the difference of two ROC curves and the difference of two AUC's. Moreover, the difference of two volumes under surfaces (VUS) is investigated to compare two treatments for the discrimination of three-category classification data, extending the ROC curve to the ROC surface in the three-dimensional case.

A few scientific progresses have been achieved in ROC curves and its related fields over the past decades. In this dissertation, we propose a plug-in empirical likelihood (EL) procedure combining placement values and weighting of inverse probability techniques, to construct stable and precise confidence intervals of the ROC curves, the difference of two ROC curves, the AUC's and the difference of two AUC's with right censoring. We proved that the limiting distribution of the EL ratio is a weighted χ^2 distribution. Furthermore, we introduce a jackknife empirical likelihood (JEL) procedure to explore the difference of two correlated VUS's with complete data. We proved that the limiting distribution of the proposed JEL ratio is a χ^2 distribution, i.e., the Wilk's theorem holds. Extensive simulation studies demonstrate that the new methods have better performance than the existing methods in terms of coverage probability of confidence intervals in most cases. Finally, the proposed methods are applied to analyze data sets of Primary Biliary Cirrhosis (PBC), Alzheimer's disease, etc.

INDEX WORDS: Area under an ROC curve, Empirical Likelihood, Jackknife empirical likelihood, Receiver operating characteristic curve, Right censored data, Volume under an ROC surface.

NOVEL NONPARAMETRIC METHODS FOR ROC CURVES

by

YUEHENG AN

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2016

NOVEL NONPARAMETRIC METHODS FOR ROC CURVES

by

YUEHENG AN

Committee Chair: Yichuan Zhao

Committee: Xin Qi

Jing Zhang

Yanqing Zhang

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

Dec 2016

DEDICATION

This dissertation is dedicated to Georgia State University.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to those who have assisted me in my graduate study. This dissertation work would not have been possible without the support of many people. I gratefully acknowledge the following people who have made the completion of this dissertation possible.

First and foremost, I owe my greatest debt of gratitude to my advisor, Dr. Yichuan Zhao, for all his endless guidance, patience, tremendous support, and for providing me with an excellent atmosphere for conducting cutting-edge research. I have learned a lot from his informative suggestions and invaluable passion on the dissertation, as well as his classes, such as Monte Carlo Methods, Statistical Inference II and Topics in Biostatistics, etc. Without his numerous careful supervision and dedicated help, I cannot finish the dissertation. I will miss his supportive advice and our discussions forever.

To my dissertation committee, Dr. Jing Zhang, Dr. Yanqing Zhang, and Dr. Xin Qi for their encouraging words, thoughtful criticism, and attention during busy semesters, and for taking their valuable time to read my dissertation and for providing me with precious suggestions. Their insightful suggestions improve my dissertation significantly.

To my professors for showing me by examples and through challenging coursework how to think, teach, and teach others to think. I would like to acknowledge all the current and previous professors who taught me throughout my graduate study in the Department of Mathematics and Statistics at Georgia State University. They immensely assisted me to enrich and broaden my statistical knowledge and skills.

To my classmates for sharing their enthusiasm and comments on my work, and for their useful help and genuine friendship.

Last but not least, I must give special thanks to my family who have always been there for me with unconditional love, understanding and encouragement.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

- AD - Alzheimer's Disease
- ADRC - Alzheimers Disease Research Center
- AL - Average length
- ANOVA - Analysis of variance
- AUC - Area under a curve
- CDF - Cumulative distribution function
- CDR - Clinical dementia rating
- CI - Confidence interval
- CLT - Central limit theorem
- CP - Coverage probability
- ECDF - Empirical cumulative distribution function
- EL - Empirical likelihood
- FNR - False negative rate
- FPR - False positive rate
- JEL - Jackknife empirical likelihood
- KM - Kaplan-Meier estimator
- LLN - Law of large numbers
- MOBVE - Marshall-Olkin bivariate exponential distribution

- NA - Normal approximation
- NPMLE - Nonparametric maximum likelihood estimate
- PBC - Primary Biliary Cirrhosis
- PVUS - Partial volume under surface
- ROC curve - Receiver operating characteristic curve
- r.v. - Random variable
- SEL - Smoothed empirical likelihood
- SLLN - Strong law of large numbers
- TNR - True negative rate
- TPR - True positive rate
- VUS - Volume under a surface
- WAIS - Wechsler Adult Intelligence Scale

CHAPTER 1

INTRODUCTION

1.1 Receiver Operating Characteristic Curve

In statistical research, a critical goal related to diagnostic medicine is to estimate and compare the accuracies of diagnostic systems. With accurate diagnostic systems, we will be able to provide reliable information about a patient's condition. Therefore, we can improve the patient care. The receiver operating characteristic curve (ROC) has been extensively applied in epidemiology, medical research, industrial quality control, signal detection, diagnostic medicine and material testing, etc. As a popular statistical tool, ROC analysis has been successfully discussed in Zweig and Campbell (1993), Metz et al. (1998b), Obuchowski (2003), Fawcett (2006), Davis and Goadrich (2006), Cook (2007), Zhou et al. (2009) and Bi et al. (2012), etc.

ROC analysis is a part of "Signal Detection Theory" developed during World War II for the analysis of radar images. Radar operators determined if a blip on the screen was a signal of a friendly ship, an enemy target or noise, etc. The ROC curve measures the ability of radar operators to make these important distinctions. In 1970's, the signal detection theory was considered as useful for interpreting medical test results.

The ROC curve describes the performance of a binary classifier system for its various discrimination thresholds. In Figure 1.1, two bell shape curves represent two populations of interest. Suppose the one on the right represents a population with disease, and the left one shows the population without disease. If a medical test value is positive, then the object is diseased. Alternatively, the test result would be negative if the test value is below the threshold. Therefore, the object is non-diseased. Sensitivity is defined as the probability of a positive test result among the population of disease, which is also called true positive rate, or TPR. Specificity is the probability of a negative test result among those without disease

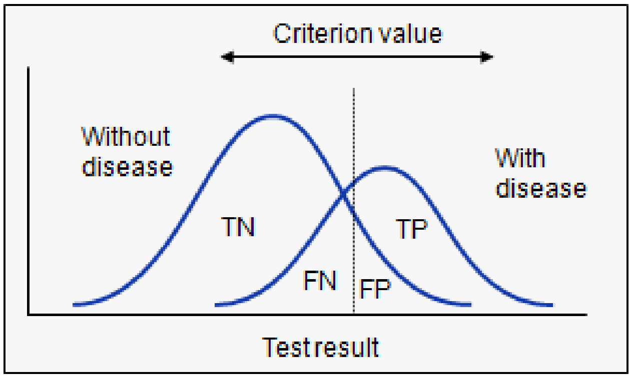

Figure 1.1: Discrimination.

referred to as true negative rate, or TNR, correspondingly. The ROC curve is a function of the sensitivity and the specificity for a measure or a model.

Let T be a continuous measurement of the results in a medical test. The disease is diagnosed if $T > t$, for a given threshold t. Denote D as the disease status with

$$D = \begin{cases} 1, & \text{diseased,} \\ 0, & \text{non-diseased,} \end{cases}$$

and let the corresponding true positive rate and false positive rate at t be $TPR(t)$ and $FPR(t)$, respectively:

$$TPR(t) = Pr(T \geq t|diseased) = Pr(T \geq t|D = 1) = sensitivity = Se,$$

and

$$FPR(t) = Pr(T \geq t|non - diseased) = Pr(T \geq t|D = 0) = 1 - specificity = 1 - Sp.$$

The ROC curve is the entire set of possible true and false positive rates attained by dichotomizing T with different thresholds (see Pepe (2003) and Zhou et al. (2009)). That is, the ROC curve is

$$ROC(\cdot) = \{(\theta, \kappa) : (FPR(t) = \theta, TPR(t) = \kappa), \, t \in (-\infty, \infty)\}.$$

Both $FPR(t)$ and $TPR(t)$ decrease when t increases. Therefore, the ROC curve is a monotonically increasing function in the positive quadrant as in Figure 1.2.

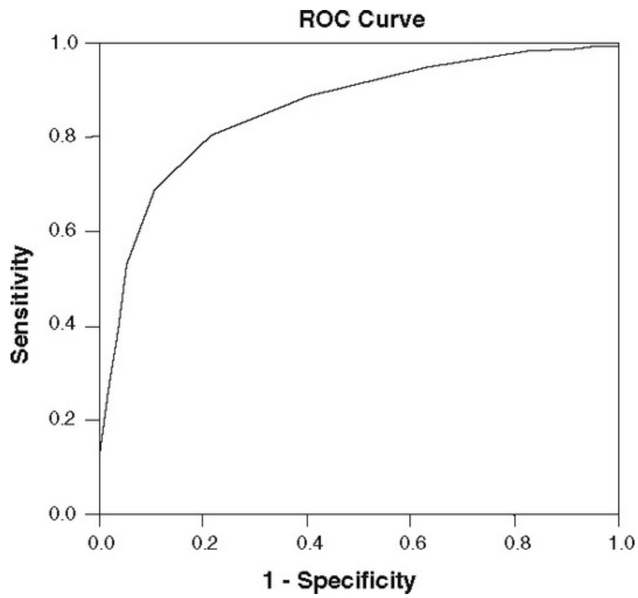

Figure 1.2: ROC curve.

Assume the distribution function of T is $F(t)$ conditional on non-disease and $G(t)$ conditional on disease. The ROC curve is defined as the graph of $(1 - G(t), 1 - F(t))$ for various values of the threshold t, where is *sensitivity* versus $(1 - specificity)$, for a test with its critical region $\{T > t\}$. At a given level $q = (1 - specificity)$, the ROC curve can be rewritten as

$$\Delta = 1 - G(F^{-1}(1 - q)), \quad for \; 0 < q < 1,$$

where F^{-1} is the inverse function of F, i.e., $F^{-1}(q) = inf\{t : F(t) \geq q\}$.

1.2 Area under an ROC Curve and Volume under a Surface

The area under the ROC curve, abbreviated AUC, provides a scalar value to summarize the performance of the learning algorithms and to compare two ROC curves in the entire range. Popular machine learning algorithms using AUC's have been found to exhibit several desirable properties when compared to accuracy, a common summary measure of medical tests (Bradley (1997)). For example, AUC has increased sensitivity in Analysis of Variance (ANOVA) tests, which is independent to the decision threshold and is invariant to a priori class of probability distributions.

In order to justify the effect of a new medicine or a new cure, physicians and medical researchers impose significant concentrations on the comparison of two treatments in clinical trials and related medical studies. A critical goal of statistical research related to diagnostic medicine is to estimate and to compare the accuracies of diagnostic systems.

As Ling et al. (2003) discussed that an AUC is a better measure than accuracy, we can choose classifiers with better AUC's, and produce better rankings. Also, Ling and Zhang (2002) showed that such classifiers produce not only better AUC's , but also better accuracy, compared to classifiers that optimize the accuracy.

Another effective method of evaluating the difference between the diagnostic accuracy of two tests is to compare areas under the receiver operating characteristic curves (AUC's). The applications of diagnostic statistical methods will help people to choose the most reliable diagnostic systems and forecast the survival times of patients with their profiles. Recent interesting literatures include Endrenyi et al. (1991), Lin et al. (1993), Hand and Till (2001), Dodd and Pepe (2003), Pencina et al. (2008), Lobo et al. (2008), Kurtcephe and Guvenir (2013), and Yang et al. (2017), among others.

The AUC's measure discrimination, i.e., the ability of the test to correctly classify those with and without the disease. Consider the situation in which patients are already correctly classified into two groups. We randomly pick one from the diseased group and one from the non-diseased group and test both. The patient with the more abnormal test result should be the one from the diseased group. The AUC is the percentage of randomly drawn pairs for which this is true, that is, the test correctly classifies the two patients of the random pair. As in Figure 1.3, AUC_B represents a test that performs better than that for AUC_A, and AUC_C is theoretically for the best test of the three AUC's.

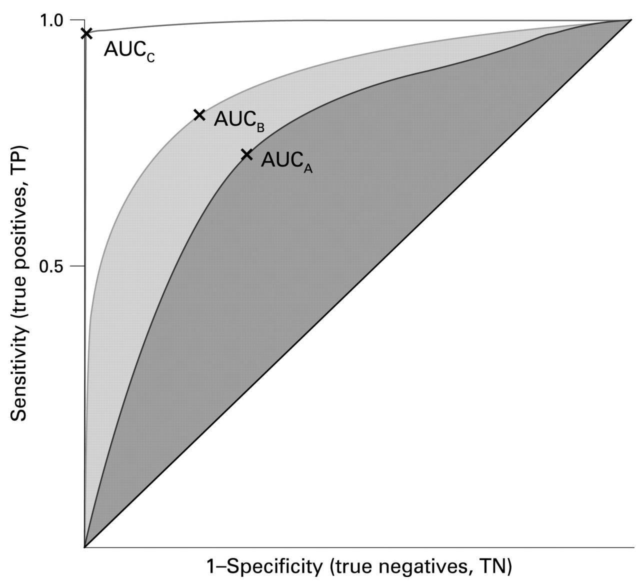

Figure 1.3: area under a ROC curve.

A multi-category classification technique is necessary if the subjects are supposed to be assigned to more than two groups simultaneously. A three-category classification treatment, for instance, can be evaluated by the volume under the ROC surface (VUS), according to Mossman (1999), Nakas and Yiannoutsos (2004), Wan (2012), etc. It is proposed as a similar measure to the AUC, extending the ROC curve to the ROC surface in a three dimensional case.

1.3 Right Censored Data

In clinical studies, the occurrence of incomplete data is common. One of the circumstances, censoring, occurs when a value occurs outside the range of a measuring instrument. The models based on censored data have many applications in medical areas, such as heart attack, cancer and HIV researches, etc. They are also widely used in engineering reliability, actuarial science, economics, finance, among others.

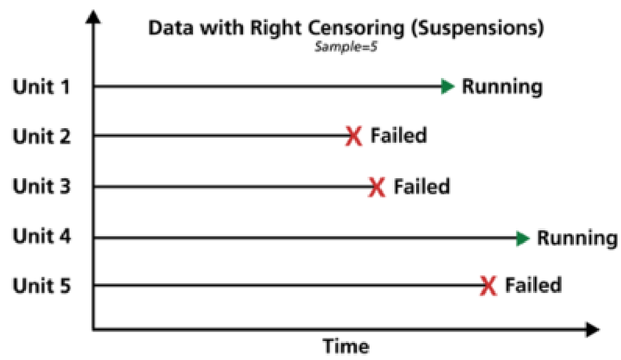

Figure 1.4: Right censored data.

A right censored value is one that is known only to be more than some value. Right censoring occurs when a subject leaves the study before an event occurs, or the study ends before the event has occurred. For example, we consider patients in a clinical trial to study the effect of treatments on stroke occurrence, and the study ends after 5 years. Those patients who have had no strokes by the end of the year are right censored. If the patient leaves the study at time t_e, then the event occurs in (t_e, ∞). Recent related work can be found in Lin and Ying (1993), Stute and Wang (1993), Stute (1995), Wang et al. (2009), Yang and Zhao (2012), and Bai and Zhou (2013), etc.

1.4 A Brief Review of Empirical Likelihood (EL) and Jackknife EL

Empirical likelihood, a nonparametric method of statistical inference, uses likelihood methods without having to assume that the data come from a known family of distributions. In other words, the empirical likelihood is a likelihood without parametric assumptions and a bootstrap without re-sampling. The approach was established by Owen in 1990's, and related work can be found in Gine and Zinn (1984), Owen (1990, 1998, 2001), Hjort et al. (2009), and Yang and Zhao (2012, 2013, 2015).

The empirical likelihood method possesses many advantages over competitors. The most appealing feature is an increase of accuracy in coverage, resulting from using easy implementation and auxiliary information. Also, the shape of the confidence regions reflects the distribution. The side information through constraints or prior distributions are incorporated straightforwardly using the empirical likelihood method.

However, empirical likelihood looses its computational advantage in applications involving nonlinear statistics, especially when the Lagrange multiplier problem reduces to solving a large number of simultaneous equations. U-statistics, for instance, exponentially aggravate the computational load when empirical likelihood is applied directly to them. Jing et al. (2009) introduce the jackknife empirical likelihood (JEL) method. It effectively resolves the computational difficulties in one and two-sample U-statistics, and applies for other nonlinear statistics as well. Also see Tian et al. (2011), Wan (2012), Kurtcephe and Guvenir (2013), and Yang and Zhao (2012, 2013, 2015), among others.

1.5 Structure

This dissertation is organized as follows. In Chapter 2, we construct the confidence intervals for ROC curves with right censored data using the empirical likelihood method. We prove that the limiting distribution of the empirical log-likelihood ratio statistic is a weighted χ^2-distribution. Then we report the results of our simulation study on the finite sample performance of the empirical likelihood based confidence intervals. Compared with the normal approximation based confidence intervals, the empirical likelihood based confidence intervals provide shorter average lengths and more precise coverage probability. In Chapter 3, a similar procedure is conducted on the difference of two ROC curves with right censored data. In Chapter 4 and Chapter 5, we move to the empirical likelihood inference on AUC's and the difference of two AUC's with right censored data. We give the asymptotic distributions of the corresponding statistics. We conduct simulation studies and report the results and corresponding conclusions. In Chapter 6, we explore the jackknife empirical likelihood confidence intervals for the difference of two volumes under the ROC surfaces with complete data. We prove that the limiting distribution of the empirical log-likelihood ratio statistic follows a χ^2-distribution. The proposed method is supported by our intensive simulation studies as well as real applications. In addition, at the end of Chapter 2 through Chapter 6, the proposed methods are applied to analyze data sets of Primary Biliary Cirrhosis (PBC), Alzheimer's disease, etc. In Chapter 7, we conclude that empirical likelihood method on the topics above outperforms the normal approximation method theoretically and practically. All the proofs are provided in the appendices.

CHAPTER 2

EMPIRICAL LIKELIHOOD FOR ROC CURVES WITH RIGHT CENSORING

2.1 Background

The receiver operating characteristic (ROC) curve is a plot of *sensitivity* versus *1-specificity* for all possible cut-off points. It provides a summary of sensitivity and specificity across a range of cut-off points for a continuous predictor. Therefore, it offers a graphical summary of the discriminatory accuracy of the diagnostic test. The ROC curve is a good statistical tool in evaluating the accuracy of tests with two-category classification data in diagnostic medicine, epidemiology, industrial quality control, and material testing, among others.

An excellent summary of recent studies is provided by Pepe (2003) and Zhou et al. (2009). Claeskens et al. (2003) has developed smoothed empirical likelihood confidence intervals for continuous-scale ROC curves with censored data. Recent interesting research work can be found in Swets and Pickett (1982), Tosteson and Begg (1988), Hsieh and Turnbull (1996), Zou et al. (1997), Lloyd (1998), Pepe (1997), Metz et al. (1998a), Lloyd and Yong (1999), Yang and Zhao (2015), etc.

The rest of this chapter is organized as follows. In Section 2.2, we construct the empirical likelihood based confidence intervals for ROC curves with right censoring for various cut-off points. The limiting distribution of the log-empirical likelihood ratio is proved to be a weighted χ^2 distribution as in Theorem 2.1. Section 2.3 reports the results of a simulation study on the finite sample performance of the empirical likelihood based confidence intervals, which outperform the confidence intervals based on the normal approximation method in terms of average length and coverage probability. In Section 2.4, we applied the proposed method to a PBC data set. All the proofs are provided in Appendix A.

2.2 Main Results

2.2.1 ROC Curves with Censoring

Let X and Y represent the populations of non-diseased and diseased patients. Let X^0 and Y^0 be the results of a continuous-scale test for the non-diseased and diseased subjects, respectively. Let F and G be the distribution functions of X^0 and Y^0. The ROC curve is defined as

$$R(p) = 1 - G(F^{-1}(1-p)), 0 < p < 1.$$

We use the same notations like Wang et al. (2009). Let $X_1^0, X_2^0, ..., X_n^0$, and $Y_1^0, Y_2^0, ..., Y_m^0$ be the random samples with distribution functions F and G. Two censoring times are $U_1, U_2, ..., U_n$, and $V_1, V_2, ..., V_m$ with distribution functions K and Q and their survival functions are $H = 1 - K$ and $L = 1 - Q$. Rather than observing X_i^0's and Y_j^0's directly, we observe (X_i, ξ_i), $i = 1, 2, ..., n$ and (Y_j, η_j), $j = 1, 2, ...m$ only, where

$$X_i = \min(X_i^0, U_i),\ \xi_i = I(X_i^0 \leq U_i),$$

$$Y_j = \min(Y_j^0, V_j),\ \eta_j = I(Y_j^0 \leq V_j),$$

where $I(\cdot)$ denotes the indicator function.

X_i^0, U_i, Y_j^0, V_j are assumed mutually independent in this dissertation, where $i = 1, 2, ..., n$ and $j = 1, 2, ...m$. We denote $\tau_f = \inf\{t : f(t) = 1\}$ for the function f. Throughout this dissertation, we assume $\tau_F \leq \tau_K$ and $\tau_G \leq \tau_Q$, and without loss of generality, we assume $\tau_F \leq \tau_G$.

2.2.2 Empirical Likelihood with Censoring

Pepe (2003) and Pepe and Cai (2004) defined the placement value as $U = 1 - F(Y^0)$. Since the CDF of the placement value

$$E(I(U \leq p)) = P(1 - F(Y) \leq p) = P(Y \geq F^{-1}(1-p)) = R(p),$$

the ROC curve can be interpreted as the distribution function of U as well. Based on the weighting of inverse probability, one has

$$E\frac{I(1-F(Y)\leq p)\eta}{1-Q(Y)} = ROC(p).$$

That is,

$$E\frac{[I(1-F(Y)\leq p)-ROC(p)]\eta}{1-Q(Y)} = 0.$$

Now we define empirical likelihood ratio for the ROC curve $ROC(p)$. Let

$$w_j = \frac{F(Y_j)-ROC(p)}{1-Q(Y_j)}\eta_j,$$

then

$$R(ROC(p)) = \sup\left\{\prod_{j=1}^{m}(mp_j), \sum_{j=1}^{m} p_j = 1, p_j > 0, \sum_{j=1}^{m} w_j p_j = 0\right\}.$$

Since F is unknown, we use the Kaplan-Meier estimator $\hat{F}$ to estimate it.

$$1-\hat{F}(t) = \prod_{X_{(i)}\leq t}(\frac{n-i}{n-i+1})^{\xi_{(i)}},$$

where $X_{(i)}$ is the ith order statistic of X-sample, that is, $X_{(1)} \leq X_{(2)} \leq ... \leq X_{(n)}$. $\xi_{(i)}$ is the corresponding ξ associated with $X_{(i)}$.Then we have

$$\hat{w}_j = \frac{[I(1-\hat{F}(Y_j)\leq p)-R(p)]\eta_j}{1-\hat{Q}(Y_j)},$$

where

$$1-\hat{Q}(t) = \prod_{Y_{(j)}\leq t}(\frac{m-j}{m-j+1})^{1-\eta_{(j)}}.$$

Then, the estimated empirical likelihood ratio for $ROC(p) = R(p)$ is

$$\hat{R}(R(p)) = \hat{R}(ROC(p)) = \sup\left\{\prod_{j=1}^{m}(mp_j), \sum_{j=1}^{m} p_j = 1, p_j > 0, \sum_{j=1}^{m} \hat{w}_j p_j = 0\right\}.$$

By the method of Lagrange multipliers, we have

$$\hat{l}(R(p)) = -2\log \hat{R}(R(p)) = 2\sum_{j=1}^{m} \log(1+\lambda \hat{w}_j),$$

where λ satisfies

$$\frac{1}{m}\sum_{j=1}^{m} \frac{\hat{w}_j}{1+\lambda \hat{w}_j} = 0.$$

Theorem 2.1. *Let $R_0(p)$ be the true value of $R(p)$. If the density functions $f(x)$ of F and $g(x)$ of G are continuous at $x = \theta_p := F^{-1}(1-p)$; $f'(x)$ and $g'(x)$ are continuous at $x = \theta_p$ and $\lim_{n\to\infty} n/m = \rho > 0$, we have*

$$\hat{l}(R(p)) \xrightarrow{\mathcal{D}} \gamma(R_0(p))\chi_1^2,$$

where the scaled constant $\gamma(R_0(p)) = \dfrac{\sigma^2(p)}{\sigma_1^2(p)}$, and

$$\sigma_1^2 = \int_0^\infty \frac{I(1-F(t)\le p) - 2R(p)I(1-F(t)\le p) + R^2(p)}{1-Q(t)} dG(t),$$

$$\sigma^2 = R_0^2(p)\sigma_y^2 + (R_0'(p))^2 p^2 \rho \sigma_x^2,$$

$$\sigma_y^2 = \int_0^{\theta_p} \frac{dG(s)}{(1-G(s))^2 L(s)},$$

$$\sigma_x^2 = \int_0^{\theta_p} \frac{dF(s)}{(1-F(s))^2 H(s)}.$$

Using Theorem 2.1, we obtain $100(1-\alpha)\%$ EL confidence intervals for $\Delta = R(p)$ as follows,

$$I_2 = \{\Delta : \hat{l}(\Delta) \le \hat{\gamma}\chi_1^2(\alpha)\},$$

where $\hat{\gamma}$ is obtained from γ by replacing the corresponding estimators, $\hat{\gamma} = \dfrac{\hat{\sigma}^2}{\hat{\sigma}_1^2}$, and

$$\hat{\sigma}_1^2 = \frac{1}{m}\sum_{j=1}^{m} \hat{w}_j^2 \xrightarrow{\mathcal{P}} \sigma_1^2 = \int_0^\infty \frac{(I(1-F(t)\le p) - 2R(p)I(1-F(t)\le p) + R^2(p)}{1-Q(t)} dG(t).$$

The consistent estimator $\hat{\sigma}^2$ of σ^2 is obtained by replacing τ_F, F, G, H, L, by $X_{(n)}$, $\hat{F}_n$, $\hat{G}_m$, $\hat{H}_n$, and $\hat{L}_m$, as defined in Section 2.2.1. Please see Appendix A for the proof in details.

2.3 Simulation Study

For the empirical likelihood procedures, the confidence intervals are given by the asymptotic distributions of the empirical log-likelihood ratios. We also implement the existing normal approximation (NA) method for the ROC curves in comparison with the empirical likelihood method, in order to justify the advantage of our proposed method. For NA method, please check the following result in Yang and Zhao (2012), Bai and Zhou (2013),

$$\sqrt{m+n}[\hat{R}(p) - R_0(p)] \xrightarrow{\mathcal{D}} N(0, \sigma^2(p)).$$

In the simulation study, the diseased population X is distributed as the exponential distribution with $\lambda_1 = 2$, while the non-diseased population Y follows the exponential distribution with $\lambda_2 = 4$. Random samples x and y are independently drawn from the population X and Y. The censoring rates for x and y are chosen as $(c_1, c_2) = (0.1, 0.1)$ and $(0.2, 0.2)$, combined with the sample sizes for x and y of $(m, n) = (50, 50)$, $(100, 100)$, $(150, 150)$. For a certain response rate and certain sample sizes, 1000 independent random samples of data $\{(x_i, \delta_{x_i}), i = 1, ..., m;\ (y_j, \delta_{y_j}), j = 1, ..., n\}$ are generated. Without loss of generality, the proposed empirical likelihood confidence intervals are constructed for the ROC curve at $q = 0.3$, 0.5, 0.7 and 0.9. The nominal levels of the confidence intervals are $1 - \alpha = 95\%$ and $1 - \alpha = 90\%$.

From Tables 2.1 - 2.4, we make the following conclusions.

1. For each censoring rate and sample size, the coverage probability is close to the nominal level, and the average lengths of the empirical likelihood based confidence intervals are shorter those based on normal approximation method;

2. In almost all the scenarios, as the censoring rates decrease or the sample sizes increase, the coverage probabilities get closer to the nominal level, and the average lengths of the

intervals decreases respectively. This is reasonable since either smaller censoring rates or bigger sample sizes provide more information of the data under study. For all different sample sizes, empirical likelihood bands are more stable and more consistent overall;

3. Empirical likelihood based confidence intervals outperform the normal approximation method.

In summary, simulation studies show that the empirical likelihood based confidence intervals outperform the normal approximation confidence intervals for small sample sizes in the sense that they yield closer coverage probabilities to the given nominal levels.

2.4 Real Application

The proposed EL method is illustrated by a data set of patients with primary biliary cirrhosis (PBC), a fatal chronic liver disease. The database is developed by the Mayo Clinic, and Fleming and Harrington (1991) has tabulated it in Appendix D.1 of their book. This randomized clinical trial includes 312 patients, 158 of whom received D-penicillamine and 154 received placebo. Among all the patients, 187 of them are censored. The censoring rate of the study is very heavy.

In this section, we construct 95% confidence intervals for the ROC curve which separates the treatment population from the placebo population. The above empirical likelihood method is implemented. We set the $(1 - specificity)$ varies from 0.01 to 0.99 by 0.01. Also, we utilize the bootstrap method to improve the accuracy, and $B = 400$ in our real data analysis. For comparison purposes, the NA method is also implemented to constructed confidence intervals for each sensitivity. In Figure 2.1, the EL confidence intervals are thinner than the NA confidence intervals, which implies that our EL method outperforms the NA method overall.

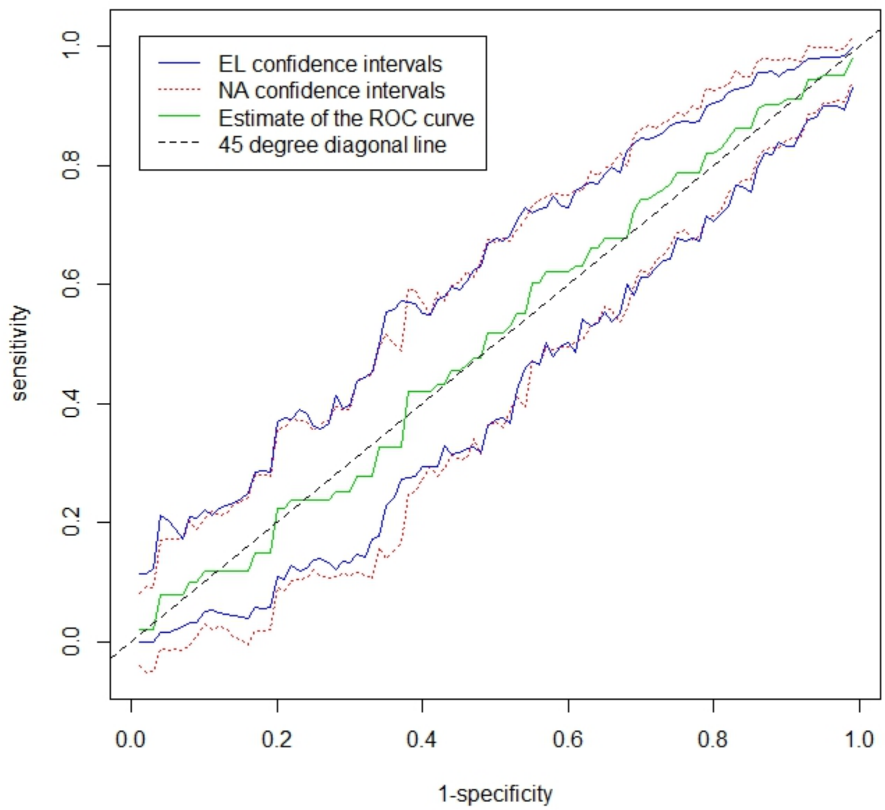

Figure 2.1: The ROC curve for the PBC data.

Table 2.1: Empirical likelihood confidence intervals for the ROC curves with right censored data at the nominal level of $1-\alpha=95\%$, X $\sim$ exp(2), Y $\sim$ exp(4), and the censoring rates $c_1=c_2=0.1$.

(m,n)	q	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	0.3	97.0	0.252	99.0	0.349
	0.5	93.4	0.370	94.8	0.392
	0.7	95.2	0.439	88.9	0.382
	0.9	95.1	0.352	84.6	0.262
(100, 100)	0.3	96.2	0.173	98.6	0.242
	0.5	94.7	0.272	94.9	0.274
	0.7	94.0	0.320	89.9	0.269
	0.9	93.6	0.265	85.1	0.192
(150, 150)	0.3	94.7	0.141	98.7	0.196
	0.5	95.3	0.221	94.5	0.223
	0.7	94.7	0.264	88.3	0.219
	0.9	95.1	0.216	84.1	0.155

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 2.2: Empirical likelihood confidence intervals for the ROC curves with right censored data at the nominal level of $1 - \alpha = 90\%$, X $\sim$ exp(2), Y $\sim$ exp(4), and the censoring rates $c_1 = c_2 = 0.1$.

(m, n)	q	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	0.3	93.1	0.212	96.8	0.293
	0.5	88.0	0.325	90.3	0.329
	0.7	89.7	0.376	80.6	0.321
	0.9	90.4	0.299	76.9	0.220
(100, 100)	0.3	91.9	0.145	96.8	0.203
	0.5	90.5	0.230	90.0	0.230
	0.7	90.3	0.270	81.8	0.225
	0.9	89.5	0.224	78.4	0.161
(150, 150)	0.3	89.8	0.118	97.9	0.164
	0.5	91.0	0.187	89.3	0.187
	0.7	90.9	0.223	80.6	0.184
	0.9	90.7	0.182	74.7	0.130

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 2.3: Empirical likelihood confidence intervals for the ROC curves with right censored data at the nominal level of $1-\alpha = 95\%$, X $\sim$ exp(2), Y $\sim$ exp(4), and the censoring rates $c_1 = c_2 = 0.2$.

(m, n)	q	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	0.3	96.5	0.276	97.2	0.397
	0.5	94.6	0.383	95.0	0.415
	0.7	95.3	0.450	87.4	0.388
	0.9	94.2	0.359	83.2	0.259
(100, 100)	0.3	95.3	0.190	98.4	0.285
	0.5	95.3	0.282	95.4	0.291
	0.7	94.2	0.325	87.6	0.273
	0.9	93.8	0.267	82.6	0.188
(150, 150)	0.3	95.7	0.156	96.8	0.223
	0.5	95.0	0.231	89.0	0.198
	0.7	95.6	0.270	88.6	0.224
	0.9	95.3	0.219	83.2	0.153

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 2.4: Empirical likelihood confidence intervals for the ROC curves with right censored data at the nominal level of $1-\alpha = 90\%$, X $\sim$ exp(2), Y $\sim$ exp(4), and the censoring rates $c_1 = c_2 = 0.2$.

(m, n)	q	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	0.3	93.4	0.233	95.5	0.333
	0.5	88.7	0.327	90.9	0.348
	0.7	90.8	0.385	80.7	0.326
	0.9	89.7	0.304	76.1	0.218
(100, 100)	0.3	90.9	0.159	96.8	0.239
	0.5	89.5	0.238	91.1	0.244
	0.7	89.6	0.276	80.2	0.229
	0.9	89.1	0.225	76.1	0.158
(150, 150)	0.3	90.9	0.131	94.9	0.187
	0.5	89.6	0.195	89.0	0.198
	0.7	91.2	0.227	81.6	0.188
	0.9	89.9	0.185	74.9	0.129

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

CHAPTER 3

EMPIRICAL LIKELIHOOD METHOD FOR THE DIFFERENCE OF 2 ROC CURVES WITH RIGHT CENSORING

3.1 Background

In medical studies, comparative benefits for alternative diagnostic algorithms, diagnostic tests, or therapeutic regimens are catching the great attention of researchers. Receiver operating characteristic (ROC) curves are widely used as a popular technique for describing and comparing the performance of diagnostic technology and diagnostics. An ROC curve is used to evaluate the discrimination ability of a diagnostic test in distinguishing the diseased population from the non-diseased population, as it visualizes the decision rule at various thresholds. Moreover, many methods are established to compare correlated or independent ROC estimates.

In clinical trials, it is appealing to select a more powerful diagnostic test from another. Two criteria can be compared in the sense of the ROC curve, and it is natural to study the differences between the two correlated ROC curves. For example, a parametric model of the difference of two ROC curves was established by Hanley and McNeil (1983), and DeLong et al. (1988) provided a nonparametric method for the difference of two correlated ROC curves. After that, Linnet (1987), Wieand et al. (1989) and Venkatraman and Begg (1996) conducted the comparison of two diagnostic tests.

The remainder of this chapter is organized as follows. In Section 3.2, we construct the empirical likelihood confidence intervals for the difference of two ROC curves with right censoring. The empirical log-likelihood ratio follows a weighted χ^2 distribution asymptotically, and the empirical likelihood based confidence intervals for various cut-off points on the ROC curve are constructed. In Section 3.3 we report the results of a simulation study on the finite sample performance of the empirical likelihood based confidence intervals. In Section 3.4, we applied the method to a PBC data set. Please see the proofs in Appendix B.

3.2 Main Results

3.2.1 The Difference of Two ROC's

Let $X = (X_1, X_2)$ and $Y = (Y_1, Y_2)$ represent the populations of non-diseased and diseased patients. Here X_1 is independent of X_2, and the sample size of X_1 is equal to that of X_2. Similarly, Y_1 is independent of Y_2, and the sample size of Y_1 is equal to that of Y_2. Let $X^0 = (X_1^0, X_2^0)$ and $Y^0 = (Y_1^0, Y_2^0)$ be the test results for the non-diseased and diseased subjects, respectively. We use $(X_{11}^0, X_{21}^0), ..., (X_{1m_1}^0, X_{2m_1}^0)$ to denote the bivariate random samples of $X^0 = (X_1^0, X_2^0)$ with the distribution function $F(x_1, x_2)$, and $(Y_{11}^0, Y_{21}^0), ..., (Y_{1m_2}^0, Y_{2m_2}^0)$ to denote the bivariate random samples of $Y^0 = (Y_1^0, Y_2^0)$ with the distribution function $G(y_1, y_2)$. Censoring times are $U_{11}, U_{12}, ..., U_{1m_1}$, and $U_{21}, U_{22}, ..., U_{2m_1}$ with distribution functions K_1 and K_2 for $X = (X_1, X_2)$, and $V_{11}, V_{12}, ..., V_{1m_2}$ and $V_{21}, V_{22}, ..., V_{2m_2}$ with distribution functions Q_1 and Q_2 for $Y = (Y_1, Y_2)$, respectively.
Instead of observing (X_{1i}^0, X_{2i}^0)'s, we observe $(X_{1i}, X_{2i}, \xi_{1i}, \xi_{2i})$, $i = 1, 2, ..., m_1$, where

$$X_{1i} = \min(X_{1i}^0, U_i),\ \xi_{1i} = I(X_{1i}^0 \le U_i),$$

$$X_{2i} = \min(X_{2i}^0, U_i),\ \xi_{2i} = I(X_{2i}^0 \le U_i).$$

That is, $\xi_i = (\xi_{1i}, \xi_{2i})$ is the indicator function of censoring, $i = 1, 2, ..., m_1$. Similarly, instead of observing (Y_{1j}^0, Y_{2j}^0)'s, we observe $(Y_{1j}, Y_{2j}, \eta_{1j}, \eta_{2j})$, $j = 1, 2, ..., m_2$, where

$$Y_{1j} = \min(Y_{1j}^0, V_j),\ \eta_{1j} = I(Y_{1j}^0 \le V_j),$$

$$Y_{2j} = \min(Y_{2j}^0, V_j),\ \eta_{2j} = I(Y_{2j}^0 \le V_j),$$

That is, $\eta_j = (\eta_{1j}, \eta_{2j})$ is the indicator of censoring, $j = 1, 2, ..., m_2$.

$X_{ki}^0, U_{ki}, Y_{kj}^0, V_{kj}$ are assumed mutually independent in this dissertation, where $i =$

$1, 2, ..., m_1$, $j = 1, 2, ...m_2$, and $k = 1, 2$. We denote $\tau_{f_k} = \inf\{t : f_k(t) = 1\}$ for the function f_k's. Throughout this dissertation, we assume $\tau_{F_k} \leq \tau_{K_k}$ and $\tau_{G_k} \leq \tau_{Q_k}$, and without loss of generality, we assume $\tau_{F_k} \leq \tau_{G_k}$, where $k = 1, 2$. We define the ROC curve with respect to the first component as

$$R_1(p) = 1 - G_1(F_1^{-1}(1-p)),\ 0 < p < 1,$$

and the ROC curve with respect to the second component as

$$R_2(p) = 1 - G_2(F_2^{-1}(1-p)),\ 0 < p < 1.$$

Then the difference of ROC curves is

$$D(p) = R_1(p) - R_2(p),$$

and the nonparametric estimator of $D(p)$ is

$$\hat{D}(p) = \hat{R}_1(p) - \hat{R}_2(p),$$

where $\hat{R}_k(p) = 1 - \hat{G}_k(\hat{F}_k^{-1}(1-p))$, $0 < p < 1$, $k = 1,\ 2$, $\hat{F}_k$'s and $\hat{G}_k$'s are the Kaplan-Meier estimators of F_k's and G_k's.

3.2.2 Empirical Likelihood with Censoring

Next, the inference of empirical likelihood (EL) based confidence intervals is discussed. Pepe (2003) and Pepe and Cai (2004) defined the placement value as $U = 1 - F(Y^0)$. For the two-sample model here, we define $U_1 = 1 - F_1(Y_1^0)$, and $U_2 = 1 - F_2(Y_2^0)$. Since the expectation of the placement value is

$$\begin{aligned} E(I(U_1 \leq p)) =& P(1 - F_1(Y_1^0) \leq p) \\ =& E(I(F_1(Y_1^0) \geq 1 - p)) \\ =& E(I(Y_1^0 \geq F_1^{-1}(1 - P))) \\ =& 1 - G_1(F_1^{-1}(1 - p)) \\ =& R_1(p). \end{aligned}$$

The ROC curve with respect to the 1st component can be interpreted as the distribution function of U_1. And

$$E(I(U_2 \leq p)) = 1 - G_2(F_2^{-1}(1 - p)) = R_2(p),$$

thus the distribution function of U_2 is also the ROC curve for the 2nd component. Therefore, the difference of two ROC curves is

$$D(p) = R_1(p) - R_2(p) = E(I(U_1 \leq p)) - E(I(U_2 \leq p)).$$

Using the weighting of the inverse probability, we have the ROC curves with right censoring as

$$E\frac{I(1 - F_1(Y_1) \leq p)\eta_1}{1 - Q_1(Y_1)} = R_1(p) \Rightarrow E\frac{[I(1 - F_1(Y_1) \leq p) - R_1(p)]\eta_1}{1 - \hat{Q}_1(Y_1)} = 0,$$

and

$$E\frac{I(1 - F_2(Y_2) \leq p)\eta_2}{1 - Q_2(Y_2)} = R_2(p) \Rightarrow E\frac{[I(1 - \hat{F}_2(Y_2) \leq p) - R_2(p)]\eta_2}{1 - \hat{Q}_2(Y_2)} = 0.$$

Then, we define empirical likelihood (EL) ratio for $D(p) = R_1(p) - R_2(p)$.

$$R(D(p)) = \sup\Big\{\prod_{k=1}^{2}\prod_{j=1}^{m_k}(m_k p_{kj}) : \sum_{j=1}^{m_k} p_{kj} = 1, p_{kj} > 0, D(p) = R_1(p) - R_2(p),$$
$$\sum_{j=1}^{m_k} \frac{I(1 - F_k(Y_{kj}) \leq p)\eta_{kj}}{1 - Q_k(Y_{kj})} p_{kj} = R_k(p), k = 1, 2\Big\}.$$

Since F_k's are unknown, we use the Kaplan-Meier (K-M) estimators $\hat{F}_k$ to estimate them:

$$1 - \hat{F}_k(t) = \prod_{X_{k(i)} \leq t} (\frac{n-i}{n-i+1})^{\xi_{k(i)}},$$

where $X_{k(i)}$ is the ith order statistics of X_k-sample, that is, $X_{k(1)} \leq X_{k(2)} \leq ...X_{k(n)}$. $\xi_{k(i)}$ is the corresponding ξ_k associated with $X_{k(i)}$.

Denote

$$\hat{w}_{kj}(p) = \frac{[(I(1 - \hat{F}_k(Y_{kj})) \leq p) - R_k(p)]\eta_{kj}}{1 - \hat{Q}_k(Y_{kj})},$$

where

$$1 - \hat{Q}_k(t) = \prod_{Y_{k(j)} \leq t} (\frac{m_k - j}{m_k - j + 1})^{1 - \eta_{k(j)}}.$$

Define the estimated empirical likelihood (EL) for $D(p)$. By the Lagrange multiplier method, we have

$$\begin{aligned} l(D(p)) =& -2\log R(D(p)) \\ =& 2\Big\{\sum_{j=1}^{m_1} \log[1 + 2\lambda\hat{w}_{1j}(p)] + \sum_{j=2}^{m_2} \log[1 - 2\lambda\hat{w}_{2j}(p)]\Big\}, \end{aligned}$$

where λ, $R_1(p)$, and $R_2(p)$ are the solutions of the following three equations:

$$\begin{cases} \frac{1}{m_1}\sum_{j=1}^{m_1} \frac{\hat{w}_{1j}(p)}{1 + 2\lambda\hat{w}_{1j}(p)} = 0, \\ \frac{1}{m_2}\sum_{j=1}^{m_2} \frac{\hat{w}_{2j}(p)}{1 - 2\lambda\hat{w}_{2j}(p)} = 0, \\ D(p) = R_1(p) - R_2(p). \end{cases}$$

Theorem 3.1. *Let $D_0(p)$ be the true value of the difference of two ROC curves $D(p)$. If (X_1, Y_1) and (X_2, Y_2) satisfy the conditions in Theorem 2.1 and $\lim_{n\to\infty} \frac{m_1}{m_2} = \gamma$, we have*

$$\hat{l}(D_0(p)) \xrightarrow{\mathcal{D}} c(p)\chi_1^2,$$

where

$$c(p) = \frac{\sigma_1^2(p) + \gamma\sigma_2^2(p)}{\sigma_{1,1}^2(p) + \gamma\sigma_{1,2}^2(p)}.$$

$(\sigma_1^2(p), \sigma_{1,1}^2(p))$ *denote the same definition of* $(\sigma^2(p), \sigma_1^2(p))$ *in Theorem 2.1 for* (X_1, Y_1) *and* $(\sigma_2^2(p), \sigma_{1,2}^2(p))$ *denote the same definition of* $(\sigma^2(p), \sigma_1^2(p))$ *in Theorem 2.1 for* (X_2, Y_2).

Furthermore, we obtain $100(1-\alpha)\%$ EL confidence intervals for $\Delta = D_0(p)$ as follows,

$$I_2 = \{\Delta : \hat{l}(\Delta) \leq \hat{c}(p)\chi_1^2(\alpha)\},$$

where

$$\hat{c}(p) = \frac{\hat{\sigma}_1^2(p) + \gamma\hat{\sigma}_2^2(p)}{\hat{\sigma}_{1,1}^2(p) + \gamma\hat{\sigma}_{1,2}^2(p)},$$

and we use the same method in Chapter 2 to get the consistent estimators $\hat{\sigma}_1^2(p)$, $\hat{\sigma}_2^2(p)$, $\hat{\sigma}_{1,1}^2(p)$ and $\hat{\sigma}_{1,2}^2(p)$. Theorem 3.1 can be extended to general bivariate cases, (X_1, Y_1) and (X_2, Y_2), when X_1 and X_2 are dependent, and Y_1 and Y_2 are dependent.

3.2.3 Normal Approximation Method

Related statistical inferences can be found in pg. 7-11 of Yao (2007), and bootstrap method can be applied to construct confidence intervals for $D(p)$. Wieand et al. (1989) showed the asymptotic distribution for the complete data,

$$\sqrt{m+n}(\hat{D}_0(p) - D_0(p)) \xrightarrow{d} N(0, \sigma^2),$$

where

$$
\begin{aligned}
\sigma^2 =&\sigma_1^2 - 2\sigma_{12} + \sigma_2^2,\\
\sigma_k =&(1-\lambda)^{-1}R_k(p_0)(1-R_k(p_0)) + \lambda^{-1}(1-p_0)p_0\frac{g_k^2(F_i^{-1}(p_0))}{f_k^2(F_i^{-1}(p_0))},\ k=1,\ 2,\\
\sigma_{12} =&(1-\lambda)^{-1}[G(F_1^{-1}(p_0), F_2^{-1}(p_0)) - G_1(F_1^{-1}(p_0))G_2(F_2^{-1}(p_0))]+\\
&\lambda^{-1}[F(F_1^{-1}(p_0), F_2^{-1}(p_0)) - p_0^2]\frac{g_1(F_1^{-1}(p_0))g_2(F_2^{-1}(p_0))}{f_1(F_1^{-1}(p_0))f_2(F_2^{-1}(p_0))},\\
\lambda =&m/(m+n),
\end{aligned}
$$

where f_k's and g_k's are the density functions of F_k's and G_k's respectively, $k = 1,\ 2$.

The normal approximation (NA) based confidence intervals can be applied to construct for the difference of two ROC curves if σ^2 was estimated properly. However, the estimation of the density functions f_k's and g_k's are rather sensitive to the choice of the smoothing parameters. The situation is similar in estimating the bivariate distribution functions $F(x_1, x_2)$ and $G(y_1, y_2)$, and the estimation of $F_k^{-1}(p)$. All of these estimations are required for the estimation of σ^2 according to the formula. Therefore, bootstrap based confidence intervals are taken into consideration.

The bootstrap based method was developed in Qin and Zhou (2006). The normal approximation (NA) based confidence intervals do not need the estimation of density functions or distribution functions. Another advantage is that it is convenient to construct bootstrap based confidence intervals through computation.

3.3 Simulation Study

We conduct an extensive simulation study to evaluate the performance of the proposed empirical likelihood confidence intervals for the difference of two ROC curves with right censored data, for different censored rates, sample sizes and nominal levels.

Let the diseased population distribute as the exponential distribution with $X_1 \sim exp(4)$, and $X_2 \sim exp(3)$where X_1 is independent of X_2; while the non-diseased population follows the exponential distribution with independent $Y_1 \sim exp(2)$ and $Y_2 \sim exp(2)$. Random samples x and y are independently drawn from the populations X and Y. The censoring rates for x and y are chosen as $(c_{x_1}, c_{x_2}, c_{y_1}, c_{y_2}) = (0.1, 0.1, 0.1, 0.1)$ and $(0.2, 0.2, 0.2, 0.2)$, combined with the sample sizes for x and y of $(m_1, m_2, n_1, n_2) = (50, 50, 50, 50)$, $(100, 100, 100, 100)$, $(150, 150, 150, 150)$. For a certain censoring rate and a certain sample size, 1000 independent random samples of data $\{(x_{ki}, \delta_{x_{ki}}), i = 1, ..., m; (y_{kj}, \delta_{y_{kj}}), j = 1, ..., n, k = 1, 2\}$ are generated. Without loss of generality, the proposed empirical likelihood confidence intervals are constructed for the ROC curve at $q = 0.1$, 0.3, 0.5, 0.7, and 0.9. The nominal levels of the confidence intervals are $1 - \alpha = 95\%$ and $1 - \alpha = 90\%$.

From Tables 3.1 - 3.4, we have the following results of the simulation study:

1. For each censoring rate, sample size and nominal levels, the coverage probability is close to the nominal level, and the average lengths of the empirical likelihood based confidence intervals are short;

2. In almost all the scenarios, as the censoring rates decrease or the sample sizes increase, the coverage probabilities get closer to the nominal level, and the average lengths of the intervals decreases respectively. This is reasonable since either smaller censoring rates or bigger sample sizes provide more information of the data under the study.

3.4 Real Application

In the previous chapter, we have discussed the PBC data about the efficacy of D-penicillamine. Moreover, in the PBC data there are also various covariates to describe the situation of the subjects. Among the covariates, the presence of hepatomegaly, that is, having an enlarged liver is a very important indicator of patients. The researchers face the task of evaluating the specific efficacy of D-penicillamine for the symptom of hepatomegaly. One way to carry out this task is to compare the ROC curves of two groups of patients with and without hepatomegaly, where both ROC curves separate the treatment population from the placebo population.

Here we construct 95% confidence intervals of the difference of 2 ROC curves at the same specificity. We set the $(1 - specificity)$ varies from 0.07 to 0.93 by 0.01. Also, we utilize the bootstrap method to improve the accuracy, and here $B = 400$.

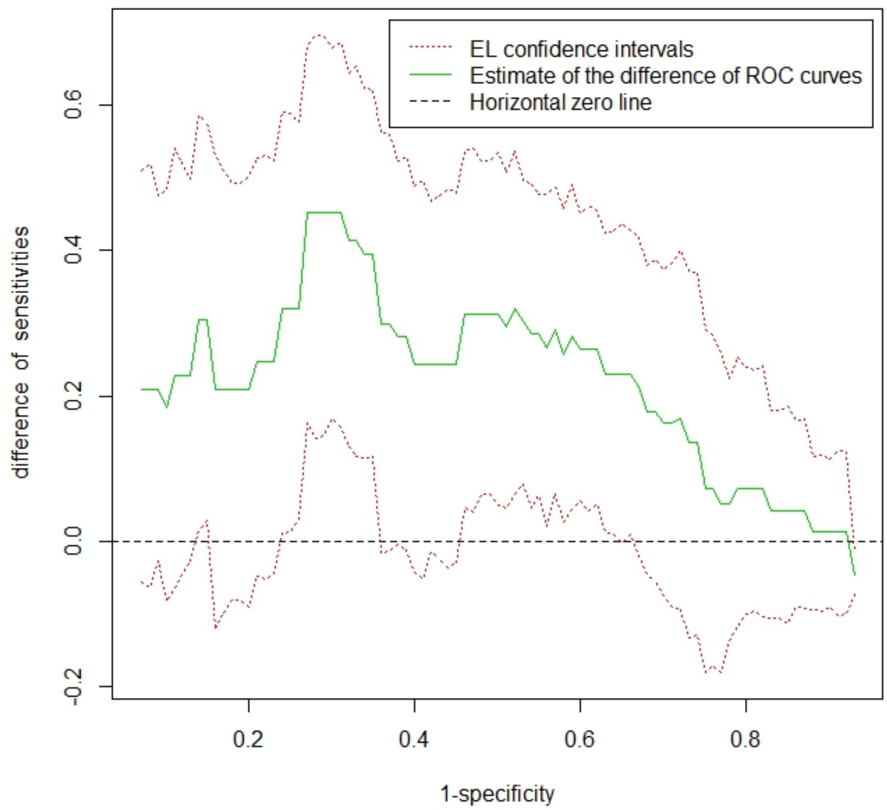

Figure 3.1: The difference of two ROC curves for the PBC data.

Figure 3.1 displays the proposed EL confidence intervals and empirical estimates for the difference of two ROC curves. Considering that the *sensitivities* of the 2 ROC curves are close at the two ends where $(1 - specificity)$ is getting close to 0 or 1, the difference of two ROC curves is close to zero. When $(1 - specificity)$ is from 0.3 to 0.7, it shows that the lower bound roughly lies above the horizontal zero line. That is, the ROC curve for the patients with hepatomegaly is higher than the ROC curve for the patients without hepatomegaly. We conclude that it is more efficient to use D-penicillamine on the patients with hepatomegaly than on the patients without hepatomegaly.

Table 3.1: Empirical likelihood confidence intervals for the difference of two ROC curves with right censoring data at the nominal level of $1-\alpha = 95\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(3)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(2)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.1$.

(m_1, m_2)	(n_1, n_2)	q	EL	
			CP(%)	AL
(50, 50)	(50, 50)	0.1	96.0	0.577
		0.3	95.4	0.552
		0.5	95.4	0.499
		0.7	94.9	0.407
		0.9	97.0	0.252
(100, 100)	(100, 100)	0.1	95.9	0.399
		0.3	94.9	0.390
		0.5	94.9	0.349
		0.7	94.9	0.289
		0.9	96.0	0.178
(150, 150)	(150, 150)	0.1	95.4	0.319
		0.3	95.0	0.325
		0.5	94.5	0.286
		0.7	95.5	0.234
		0.9	95.6	0.141

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 3.2: Empirical likelihood confidence intervals for the difference of two ROC curves with right censoring data at the nominal level of $1-\alpha = 90\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(3)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(2)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.1$.

(m_1, m_2)	(n_1, n_2)	q	EL	
			CP(%)	AL
(50, 50)	(50, 50)	0.1	91.9	0.481
		0.3	91.1	0.464
		0.5	90.4	0.419
		0.7	89.4	0.341
		0.9	92.4	0.211
(100, 100)	(100, 100)	0.1	90.6	0.333
		0.3	89.2	0.327
		0.5	89.6	0.293
		0.7	90.1	0.243
		0.9	91.8	0.149
(150, 150)	(150, 150)	0.1	90.1	0.267
		0.3	90.4	0.264
		0.5	90.2	0.240
		0.7	91.1	0.196
		0.9	90.6	0.118

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 3.3: Empirical likelihood confidence intervals for the difference of two ROC curves with right censoring data at the nominal level of $1-\alpha = 95\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(3)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(2)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.2$.

(m_1, m_2)	(n_1, n_2)	q	EL	
			CP(%)	AL
(50, 50)	(50, 50)	0.1	94.5	0.684
		0.3	94.8	0.587
		0.5	94.5	0.519
		0.7	95.5	0.422
		0.9	95.6	0.266
(100, 100)	(100, 100)	0.1	95.5	0.460
		0.3	94.0	0.414
		0.5	94.1	0.360
		0.7	94.6	0.296
		0.9	95.9	0.185
(150, 150)	(150, 150)	0.1	95.5	0.365
		0.3	95.0	0.331
		0.5	94.6	0.294
		0.7	95.3	0.239
		0.9	96.8	0.148

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 3.4: Empirical likelihood confidence intervals for the difference of two ROC curves with right censoring data at the nominal level of $1-\alpha = 90\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(3)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(2)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.2$.

(m_1, m_2)	(n_1, n_2)	q	EL	
			CP(%)	AL
(50, 50)	(50, 50)	0.1	89.4	0.577
		0.3	90.1	0.491
		0.5	90.7	0.435
		0.7	90.5	0.354
		0.9	92.5	0.223
(100, 100)	(100, 100)	0.1	91.3	0.381
		0.3	88.9	0.346
		0.5	89.0	0.302
		0.7	90.2	0.248
		0.9	90.9	0.155
(150, 150)	(150, 150)	0.1	91.1	0.303
		0.3	89.8	0.277
		0.5	89.5	0.247
		0.7	90.5	0.200
		0.9	91.6	0.124

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

CHAPTER 4

EMPIRICAL LIKELIHOOD FOR THE AREA UNDER THE ROC CURVE WITH RIGHT CENSORING

4.1 Background

Accurate diagnostic systems can provide reliable information about a patient's condition and improve patient care. Sometimes, an ROC curve is not very convenient as a two-dimensional depiction of the classification performance because each ROC curve consists a series of ordered pairs. In this case, the area under the curve (AUC) is proposed as a commonly used summary index of the ROC curve telling the overall classification performance. As the name implies, an AUC is the integral of the ROC curve at the interval $(0, 1)$. Larger AUC value indicates stronger discrimination ability. Thus, it represents a more effective treatment. The applications of the diagnostic statistical methods will help the users to choose a more reliable diagnostic system over another, and forecast the survival times of patients by looking at their profiles.

Machine learning has become more and more popular in the academia and industry as a branch of artificial intelligence in the recent years. The AUC has been proved to be a better measure than the accuracy in the meanings of consistency and discriminancy when the two evaluation measures for learning algorithms are compared. Ling et al. (2003) presented a rigourous proof as well as empirical evaluations that the area under curves (AUC) is more efficient than accuracy as a statistical measure (Valeinis (2007)).

The rest of this chapter is organized as follows. In Section 4.2, we present the confidence intervals based on the empirical likelihood method for the AUCs with right censored data, and the limiting distribution of the statistic is a weighted χ^2 distribution. In Section 4.3 we report the results of a simulation study on the finite sample performance of the empirical likelihood based confidence intervals. Compared with the normal approximation method, empirical likelihood confidence intervals outperform in terms of average length and coverage probability. In Section 4.4, we applied the proposed method to a PBC data set. All the proofs are included in Appendix C.

4.2 Main Results

4.2.1 Area under the ROC Curves with Censoring

Let X and Y represent the populations of non-diseased and diseased patients. Let X^0 and Y^0 be the results of a continuous-scale test for the non-diseased and diseased subjects, respectively. Let F and G be the distribution functions of X^0 and Y^0. The ROC curve is defined as

$$R(p) = 1 - G(F^{-1}(1-p)), 0 < p < 1.$$

The area under the ROC curve (AUC) is defined as

$$\Delta = \int_0^1 R(p)dp.$$

4.2.2 Normal Approximation Procedure for Δ with Censoring

We use the same notations like Wang et al. (2009). Let $X_1^0, X_2^0, ..., X_n^0$, and $Y_1^0, Y_2^0, ..., Y_m^0$ be the random samples with distribution functions F and G. Two censoring times are $U_1, U_2, ..., U_n$, and $V_1, V_2, ..., V_m$ with distribution functions K and Q. Rather than observing X_i^0's and Y_j^0's directly, we were only able to observe only (X_i, ξ_i), $i = 1, 2, ..., n$ and (Y_j, η_j), $j = 1, 2, ...m$, where

$$X_i = \min(X_i^0, U_i),\ \xi_i = I(X_i^0 \leq U_i),$$

$$Y_j = \min(Y_j^0, V_j),\ \eta_j = I(Y_j^0 \leq V_j),$$

$I(\cdot)$ is the indicator function.

One supposes that X_i^0, U_i, Y_j^0, V_j are mutually independent. We denote $\tau_f = \inf\{t : f(t) = 1\}$ for the function f. In this dissertation, we assume $\tau_F \leq \tau_K$ and $\tau_G \leq \tau_Q$. Without loss of generality, we assume $\tau_F \leq \tau_G$. Then the AUC is

$$\Delta = P(X^0 < Y^0) = \int_0^{\tau_F} (1 - G(t))dF(t).$$

$F(t)$ and $G(t)$ are unknown, so we replace them with Kaplan-Meier (KM) estimates $\hat{F}(t)$ and $\hat{G}(t)$ as follows:

$$1 - \hat{F}(t) = \prod_{X_{(i)} \leq t} (\frac{n-i}{n-i+1})^{\xi_{(i)}},$$

$$1 - \hat{G}(t) = \prod_{Y_{(j)} \leq t} (\frac{m-j}{m-j+1})^{\eta_{(j)}},$$

where $X_{(i)}$ is the ith order statistics of X-sample, and $Y_{(j)}$ is the jth order statistics of Y-sample, $\xi_{(i)}$ and $\eta_{(j)}$ are the corresponding ξ and η associated with $X_{(i)}$ and $Y_{(j)}$.

Let $H(t) = P(X \leq t)$, and $L(t) = P(Y \leq t)$. Denote

$$\Lambda_F(t) = \int_0^t \frac{dF(s)}{1 - F(s-)},$$

and

$$\Lambda_G(t) = \int_0^t \frac{dG(s)}{1 - G(s-)}.$$

Moreover, let $\hat{H}_n(t) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq t)$, and $\hat{L}_m(t) = \frac{1}{m}\sum_{j=1}^{m} I(Y_j \leq t)$.

Wang et al. (2009) have shown the following theorem is true. Here we review it for completeness.

Theorem 4.1 (Wang et al. (2009))**.** *Let Δ_0 be the true value of Δ. Under the regularity conditions (1)-(5) given in Wang et al. (2009),*

(1) $n/m \to \rho$, $\rho > 0$;

(2)

(i) $\sqrt{n+m}\int_{X_{(n)}}^{\tau_F} F(t)dG(t) \xrightarrow{\mathcal{P}} 0$;

(ii) $\sqrt{n+m}(G(\tau_F) - G(X_{(n)})) \xrightarrow{\mathcal{P}} 0$;

(iii) $\sqrt{n+m}\int_{X_{(n)}}^{\tau_F}(1 - G(t))dF(t) \xrightarrow{\mathcal{P}} 0$;

(3)

(i) $sup_t|\int_t^{\tau_F}(1-F(s))dG(t)/(1-F(t))| < \infty$*;*

(ii) $sup_t|\int_t^{\tau_F}(1-G(s))dF(t)/(1-G(t))| < \infty$*;*

(4)

(i) $\int_0^{\tau_F} dF(t)/(1-K(t-)) < \infty$*;*

(ii) $\int_0^{\tau_F} dG(t)/(1-Q(t-)) < \infty$*;*

(5) N_{xi} *and* N_{yj} *have no common jumps, respectively, where* $N_{(xi)} = I(X_i \leq t, \xi_i = 1)$, $N_{(yj)} = I(Y_j \leq t, \eta_j = 1)$ *for* $i = 1, 2, ..., n,$ *and* $j = 1, 2, ..., m.$

$$\sqrt{m+n}(\hat{\Delta} - \Delta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\sigma^2 = (1 + 1/\rho)\sigma_x^2 + (1 + \rho)\sigma_y^2$, $\rho = \lim n/m$,

$$\sigma_x^2 = \int_0^{\tau_F}((1-F(t))(1-G(t)) - \int_t^{\tau_F}(1-G(s))dF(s))^2 \frac{1-F(t-)}{1-F(t)} \frac{1}{1-H(t-)} d\Lambda_F(t),$$

$$\sigma_y^2 = \int_0^{\tau_F}(\int_t^{\tau_F}(1-G(s))dF(s))^2 \frac{1-G(t-)}{1-G(t)} \frac{1}{1-L(t-)} d\Lambda_G(t).$$

The consistent estimator $\hat{\sigma}^2$ of σ^2 is obtained by replacing τ_F, F, G, H, L, by $X_{(n)}$, $\hat{F}_n$, $\hat{G}_m$, $\hat{H}_n$, and $\hat{L}_m$. Based on Theorem 4.1, the $100(1-\alpha)\%$ normal approximation confidence interval for Δ is

$$I_1 = \left\{ \Delta : |\Delta - \Delta_0| \leq Z_{\alpha/2}\sqrt{\frac{\hat{\sigma}^2}{m+n}} \right\},$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ critical value for the standard normal distribution $N(0,1)$.

4.2.3 Empirical Likelihood Procedure for Δ with Censoring

Next, we make the inference using empirical likelihood method. Pepe and Cai (2004) and Qin and Zhou (2006) defined the placement value as $P = 1 - F(Y_0)$, and $E(1-P) = E(F(Y_0)) = \Delta_0$. Based on the weighting of the inverse probability, we have

$$F(Y_0) = P(X_0 < Y_0),$$

$$\Delta_0 = E\frac{F(Y)\eta}{1-Q(Y)}.$$

Let

$$Z_j = \frac{(F(Y_j)-\Delta)\eta_j}{1-Q(Y_j)}.$$

We define empirical likelihood ratio for Δ_0 as follows:

$$R(\Delta) = \sup\left\{\prod_{j=1}^{m}(mp_j), \sum_{j=1}^{m} p_j = 1, p_j > 0, \sum_{j=1}^{m}\frac{(F(Y_j)-\Delta)\eta_j}{1-Q(Y_j)}p_j = 0\right\}.$$

Since F, Q are unknown, we use Kaplan-Meier estimator $\hat{F}$, $\hat{Q}$ to estimate them. Thus the estimated empirical likelihood ratio for Δ is as follows. Let

$$\hat{Z}_j = \frac{(\hat{F}(Y_j)-\Delta)\eta_j}{1-\hat{Q}(Y_j)},$$

where

$$1-\hat{F}(t) = \prod_{X_{(i)}\le t}\left(\frac{n-i}{n-i+1}\right)^{\xi_{(i)}},$$

$$1-\hat{Q}(t) = \prod_{Y_{(j)}\le t}\left(\frac{m-j}{m-j+1}\right)^{1-\eta_{(j)}}.$$

Therefore

$$\hat{R}(\Delta) = \sup\Big\{\prod_{j=1}^{m}(mp_j), \sum_{j=1}^{m} p_j = 1, p_j > 0, \sum_{j=1}^{m}\frac{(\hat{F}(Y_j)-\Delta)\eta_j}{1-\hat{Q}(Y_j)}p_j = 0\Big\}.$$

By the Lagrange multiplier method, we have

$$\hat{l}(\Delta) = -2\log\hat{R}(\Delta) = 2\sum_{j=1}^{m}\log\Big(1+\lambda\frac{(\hat{F}(Y_j)-\Delta)\eta_j}{1-\hat{Q}(Y_j)}\Big),$$

where λ satisfies

$$\frac{1}{m}\sum_{j=1}^{m}\frac{\hat{Z}_j}{1+\lambda\hat{Z}_j} = 0.$$

Next we state the following theorem and show how to construct confidence intervals for

Δ.

Theorem 4.2. *Under the regularity conditions (1)-(5) given by Wang et al. (2009), as in the above theorem, we have*

$$\hat{l}(\Delta_0) \xrightarrow{\mathcal{D}} \gamma(\Delta_0)\chi_1^2,$$

where $\gamma(\Delta_0) = \dfrac{\sigma^2}{\sigma_1^2}$, *and* $\sigma_1^2 = \int_0^\infty \dfrac{(F-\Delta_0)^2}{1-Q(t)} dG(t)$.

Remark 4.1. *In special cases, the data is noncensoring, or complete, then*

$$\begin{aligned}
\sigma_1^2 &= \int_0^\infty (F-\Delta_0)^2 dG(t) \\
&= \int_0^\infty (F^2 - 2\Delta_0 F + \Delta_0^2) dG(t) \\
&= \int_0^\infty F^2 dG(t) - \int_0^\infty 2\Delta_0 F dG(t) + \int_0^\infty \Delta_0^2 dG(t) \\
&= \int_0^\infty F^2 dG(t) - 2\Delta_0\Delta_0 + \Delta_0^2 \\
&= \int_0^\infty F^2 dG(t) - \Delta_0^2.
\end{aligned}$$

Theorem 4.2 can be proved using Theorem 2.1 in Hjort et al. (2009). Please check the theorem in Appendix C. Furthermore, we obtain $100(1-\alpha)\%$ EL confidence intervals for Δ as follows,

$$I_2 = \{\Delta : \hat{l}(\Delta) \leq \hat{\gamma}\chi_1^2(\alpha)\},$$

where $\hat{\gamma}$ is obtained from γ by replacing the corresponding estimators, $\hat{\gamma} = \dfrac{\hat{\sigma}^2}{\hat{\sigma}_1^2}$, and $\hat{\sigma}_1^2 = \dfrac{1}{m}\sum_{j=1}^{m} \hat{Z}_j^2$.

4.3 Simulation Study

In this section, we conduct a simulation study to investigate the finite sample performance of the proposed empirical likelihood confidence intervals for the AUCs with right censored data, for different censored rates, sample sizes, nominal levels and different parameters of exponential distributions. For comparison purposes, we also construct the confidence intervals based on NA method. See Section 4.2.2 for details.

In the simulation studies, we have the same settings as in the ROC curve with right censoring. The diseased population X is distributed as the exponential distribution with λ_1, while the non-diseased population Y follows the exponential distribution with λ_2. Various values of λ_1 and λ_2 are chosen as follows. Random samples x and y are independently drawn from the population X and Y. The censoring rates for x and y are chosen as $(c_1, c_2) = (0.2, 0.2)$, combined with the sample sizes for x and y of $(m, n) = (50, 50)$, $(100, 100)$, $(150, 150)$. For a certain response rate and a certain sample size, 1000 independent random samples of data $\{(x_i, \delta_{x_i}), i = 1, ..., m; (y_j, \delta_{y_j}), j = 1, ..., n\}$ are generated. The nominal levels of the confidence intervals are $1 - \alpha = 95\%$ and $1 - \alpha = 90\%$.

Tables 4.1 - 4.4 display the simulation results, and we make the conclusions as follows:
1. For each censor rate, sample size and parameters of the distribution, the coverage probability is close to the nominal level, and the average lengths of the confidence intervals are short;
2. In almost all the scenarios, as the sample sizes increase or censor rates decrease, the coverage probabilities get closer to the nominal level, and the average lengths of the intervals decreases respectively. This is reasonable since either larger response rates or larger sample sizes provide more information of the data under study;
3. Empirical likelihood based confidence intervals outperform the normal approximation method.

4.4 Real Application

In this section, we implement the proposed empirical likelihood method on the same PBC data for the randomized clinical trial of 312 patients, as in Chapter 2. Similar to Wang et al. (2009), we employed the estimate of AUC instead of the true value of AUC. Also, we utilize the bootstrap method to improve the accuracy, and $B = 400$ in our data analysis.

The 95% EL confidence interval is $(0.435, 0.589)$, and the 90% EL confidence interval is $(0.448, 0.576)$. For comparison purpose, the NA method is also implemented to constructed confidence intervals. The 95% NA confidence interval is $(0.406, 0.618)$, and the 90% NA confidence interval is $(0.423, 0.601)$. The two groups of confidence intervals both contain 0.5 which is consistent with the behavior of the ROC curve in Chapter 2. Meanwhile, the EL confidence intervals are narrower than the NA confidence intervals which implies that EL method will be more accurate.

Table 4.1: Empirical likelihood confidence intervals for the area under ROC curves (AUC) with right censoring data at the nominal level of $1-\alpha = 95\%$, and the censoring rates $c_1 = c_2 = 0.2$.

(m,n)	(λ_1,λ_2)	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	(2,10)	91.9	0.159	92.8	0.175
	(4, 8)	92.8	0.216	93.9	0.223
	(6, 6)	94.1	0.233	93.6	0.238
	(8, 4)	94.8	0.219	94.3	0.222
	(10, 2)	91.9	0.159	94.3	0.174
(100, 100)	(2,10)	94.4	0.097	93.6	0.120
	(4, 8)	93.9	0.155	93.6	0.156
	(6, 6)	94.2	0.166	94.4	0.167
	(8, 4)	95.7	0.155	94.1	0.155
	(10, 2)	94.1	0.118	85.1	0.119
(150, 150)	(2,10)	94.4	0.097	94.4	0.097
	(4, 8)	94.6	0.128	95.0	0.127
	(6, 6)	94.4	0.136	94.4	0.137
	(8, 4)	94.5	0.126	94.5	0.127
	(10, 2)	94.4	0.097	93.8	0.097

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 4.2: Empirical likelihood confidence intervals for the area under ROC curves (AUC) with right censoring data at the nominal level of $1-\alpha = 90\%$, and the censoring rates $c_1 = c_2 = 0.2$.

(m, n)	(λ_1, λ_2)	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	(2,10)	91.9	0.159	88.1	0.147
	(4, 8)	87.5	0.181	87.6	0.187
	(6, 6)	88.5	0.196	88.4	0.200
	(8, 4)	89.2	0.184	89.3	0.187
	(10, 2)	85.2	0.133	89.9	0.146
(100, 100)	(2,10)	88.9	0.098	89.2	0.101
	(4, 8)	89.2	0.130	88.8	0.131
	(6, 6)	90.2	0.140	88.8	0.140
	(8, 4)	90.1	0.130	89.5	0.130
	(10, 2)	88.9	0.098	89.1	0.100
(150, 150)	(2,10)	89.1	0.081	89.6	0.082
	(4, 8)	90.2	0.107	89.3	0.107
	(6, 6)	89.4	0.114	89.7	0.115
	(8, 4)	90.4	0.106	89.4	0.107
	(10, 2)	89.1	0.081	88.9	0.081

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 4.3: Empirical likelihood confidence intervals for the area under ROC curves (AUC) with right censoring data at the nominal level of $1 - \alpha = 95\%$, and the censoring rates $c_1 = c_2 = 0.1$.

(m, n)	(λ_1, λ_2)	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	(2,10)	92.7	0.161	92.5	0.164
	(4, 8)	93.8	0.212	93.6	0.216
	(6, 6)	94.5	0.227	94.1	0.232
	(8, 4)	94.9	0.212	94.1	0.216
	(10, 2)	94.4	0.164	93.4	0.165
(100, 100)	(2,10)	94.8	0.117	93.4	0.115
	(4, 8)	94.1	0.152	94.1	0.152
	(6, 6)	94.8	0.162	93.4	0.163
	(8, 4)	95.7	0.150	94.1	0.151
	(10, 2)	95.6	0.115	94.9	0.114
(150, 150)	(2,10)	94.7	0.096	94.1	0.094
	(4, 8)	94.9	0.124	95.0	0.124
	(6, 6)	95.3	0.132	95.3	0.133
	(8, 4)	95.2	0.123	94.9	0.124
	(10, 2)	94.0	0.094	93.8	0.094

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 4.4: Empirical likelihood confidence intervals for the area under ROC curves (AUC) with right censoring data at the nominal level of $1-\alpha = 90\%$, and the censoring rates $c_1 = c_2 = 0.1$.

(m, n)	(λ_1, λ_2)	EL		NA	
		CP(%)	AL	CP(%)	AL
(50, 50)	(2,10)	86.8	0.134	87.9	0.138
	(4, 8)	88.8	0.178	89.1	0.181
	(6, 6)	89.1	0.191	88.3	0.195
	(8, 4)	89.2	0.178	89.0	0.181
	(10, 2)	89.6	0.137	89.2	0.138
(100, 100)	(2,10)	89.3	0.097	88.2	0.097
	(4, 8)	89.5	0.127	88.0	0.127
	(6, 6)	89.8	0.136	89.2	0.136
	(8, 4)	91.1	0.126	88.6	0.127
	(10, 2)	90.4	0.096	88.0	0.096
(150, 150)	(2,10)	89.9	0.080	89.7	0.079
	(4, 8)	90.5	0.104	89.3	0.104
	(6, 6)	89.8	0.111	90.0	0.112
	(8, 4)	89.8	0.103	89.3	0.104
	(10, 2)	89.9	0.078	90.2	0.079

NOTE:
EL: empirical likelihood,
NA: normal approximation,
CP(%): coverage probability,
AL: average length of a confidence interval.

CHAPTER 5

EMPIRICAL LIKELIHOOD FOR THE DIFFERENCE OF 2 AUC'S WITH UNIVARIATE CENSORING

5.1 Background

Another effective method to evaluate the difference between the diagnostic accuracy of two tests is to find the difference of two AUC's. The applications of diagnostic statistical methods will help users to make a choice of the most reliable diagnostic systems and to forecast the survival times of patients with their profile. Huang et al. (2012) have studied the difference between two AUCs with complete data.

The rest of this chapter is organized as follows. In Section 5.2, we show the empirical likelihood method confidence intervals for the difference of two AUCs with right censored data and give the limiting distribution of the statistic. In Section 5.3, we display the results from a simulation study on the finite sample performance of the empirical likelihood based confidence intervals and the performance of the methods theoretically. Compared with those from the normal approximation method, the confidence intervals based on empirical likelihood method outperform in terms of average length and coverage probability. In Section 5.4, we applied the method to a PBC data set. All the proofs are provided in Appendix D.

5.2 Main Results

5.2.1 The Difference of Two AUC's

We use the same notations as in Chapter 3. Please also see Lin and Ying (1993) for the notations. We define the AUC with the first component as

$$\Delta_1 = P(X_1^0 < Y_1^0) = \int_0^{\tau_{F_1}} (1 - G_1(t))dF_1(t).$$

and the AUC with the second component as

$$\Delta_2 = P(X_2^0 < Y_2^0) = \int_0^{\tau_{F_2}} (1 - G_2(t)) dF_2(t).$$

Therefore, We define the difference of AUC's

$$\Delta = \Delta_1 - \Delta_2.$$

5.2.2 Empirical Likelihood for Δ with Censoring

The inference of empirical likelihood (EL) based confidence intervals are discussed in the following way. Pepe and Cai (2004) defined the placement value as $U = 1 - F(Y_0)$. For the two-sample data here, we define $P_k = 1 - F_k(Y_0)$, and $E(1 - P_k) = E(F_k(Y_{k0})) = \Delta_{k0}$, $k = 1, 2$. Using the weighting of the inverse probability for right censoring data, we have

$$E \frac{F_k(Y_k)\eta_k}{1 - Q_k(Y_k)} = \Delta_{k0},$$

that is,

$$E \frac{(F_k(Y_k) - \Delta_{k0})\eta_k}{1 - Q_k(Y_k)} = 0.$$

We define empirical likelihood ratio for $\Delta = \Delta_1 - \Delta_2$ as

$$R(\Delta_0) = \sup \Big\{ \prod_{k=1}^{2} \prod_{j=1}^{m_k} (m_k p_{kj}), \sum_{j=1}^{m_k} p_{kj} = 1, p_{kj} > 0, \sum_{j=1}^{m_k} \Big[\frac{F_k(Y_{kj})\eta_{kj}}{1 - Q_k(Y_{kj})} - \Delta_{k0} \Big] p_{kj} = 0, k = 1, 2,$$
$$\sum_{j=1}^{m_1} \frac{F_1(Y_{1j})\eta_{1j}}{1 - Q_1(Y_{1j})} p_{1j} - \sum_{j=1}^{m_2} \frac{F_2(Y_{2j})\eta_{2j}}{1 - Q_2(Y_{2j})} p_{2j} = \Delta_0 \Big\},$$

or

$$R(\Delta_0) = \sup \Big\{ \prod_{k=1}^{2} \prod_{j=1}^{m_k} (m_k p_{kj}), \sum_{j=1}^{m_k} p_{kj} = 1, p_{kj} > 0, \sum_{j=1}^{m_k} (w_{kj} - \Delta_{k0}) p_{kj} = 0, k = 1, 2,$$
$$\Delta_0 = \Delta_{10} - \Delta_{20} = \sum_{j=1}^{m_1} w_{1j} p_{1j} - \sum_{j=1}^{m_2} w_{2j} p_{2j} \Big\},$$

where $w_{kj} = \dfrac{F_k(Y_{kj})\eta_{kj}}{1 - Q_k(Y_{kj})}$, and $\Delta_{k0} = \sum_{j=1}^{m_k} w_{kj} p_{kj}$.

Since F_k's are unknown, the Kaplan-Meier (K-M) estimator $\hat{F}_k$ is applied to estimate them.

$$1 - \hat{F}_k(t) = \prod_{X_{k(i)} \le t} \left(\frac{n-i}{n-i+1}\right)^{\xi_{k(i)}},$$

where $X_{k(i)}$ is the ith order statistics of X_k-sample, that is, $X_{k(1)} \le X_{k(2)} \le ... \le X_{k(m_1)}$. $\xi_{k(i)}$ is the corresponding ξ_k associated with $X_{k(i)}$.

Denote

$$\hat{w}_{kj} = \frac{\hat{F}_k(Y_{kj})\eta_{kj}}{1 - \hat{Q}_k(Y_{kj})},$$

where

$$1 - \hat{Q}_k(t) = \prod_{Y_{(kj)} \le t} \left(\frac{m_k - j}{m_k - j + 1}\right)^{1 - \eta_{(kj)}}.$$

By the Lagrange multiplier method, we have

$$\begin{aligned} l(\Delta) =& -2 \log R(\Delta) \\ =& 2\Big\{ \sum_{j=1}^{m_1} \log[1 - 2\lambda(\hat{w}_{1j} - \Delta_1)] + \sum_{j=1}^{m_2} \log[1 + 2\lambda(\hat{w}_{2j} - \Delta_2)] \Big\}, \end{aligned}$$

where λ, Δ_1, and Δ_2 are the solutions to the following three equations,

$$\frac{1}{m_1} \sum_{j=1}^{m_1} \frac{\hat{w}_{1j} - \Delta_1}{1 - 2\lambda(\hat{w}_{1j} - \Delta_1)} = 0,$$

$$\frac{1}{m_2} \sum_{j=1}^{m_2} \frac{\hat{w}_{2j} - \Delta_2}{1 + 2\lambda(\hat{w}_{2j} - \Delta_2)} = 0,$$

$$\frac{1}{m_1} \sum_{j=1}^{m_1} \frac{\hat{w}_{1j}}{1 - 2\lambda(\hat{w}_{1j} - \Delta_1)} - \frac{1}{m_2} \sum_{j=1}^{m_2} \frac{\hat{w}_{2j}}{1 + 2\lambda(\hat{w}_{2j} - \Delta_2)} = \Delta_1 - \Delta_2 = \Delta.$$

Theorem 5.1. *Let $\Delta_1 - \Delta_2 = \Delta_0$ be the true value of the difference of two AUC's. If*

(X_1, Y_1) and (X_2, Y_2) satisfy the conditions in Theorem 4.1 and $\lim_{n\to\infty} \frac{m_1}{m_2} = \gamma$ *, we have*

$$\hat{l}(\Delta_0) \xrightarrow{\mathcal{D}} c(\Delta_0)\chi_1^2,$$

where

$$c(\Delta_0) = \frac{\sigma_1^2 + \gamma\sigma_2^2}{\sigma_{1,1}^2 + \gamma\sigma_{1,2}^2}.$$

$(\sigma_1^2, \sigma_{1,1}^2)$ denote the same definition of (σ^2, σ_1^2) in Theorem 4.1 for (X_1, Y_1) and $(\sigma_2^2, \sigma_{1,2}^2)$ denote the same definition of (σ^2, σ_1^2) in Theorem 4.1 for (X_2, Y_2).

Using Theorem 5.1, we obtain the asymptotic $100(1-\alpha)\%$ EL confidence interval for Δ as follows,

$$I_2(\Delta) = \{\Delta : \hat{l}(\Delta) \le \hat{c}\chi_1^2(\alpha)\},$$

where

$$\hat{c} = \frac{\hat{\sigma}_1^2 + \gamma\hat{\sigma}_2^2}{\hat{\sigma}_{1,1}^2 + \gamma\hat{\sigma}_{1,2}^2},$$

and we use the same method in Chapter 4 to get the consistent estimators $\hat{\sigma}_1^2$, $\hat{\sigma}_2^2$, $\hat{\sigma}_{1,1}^2$ and $\hat{\sigma}_{1,2}^2$.

5.3 Simulation Study

Based on the conclusions from the inference procedure, we conduct an extensive simulation study to evaluate the performance of the proposed empirical likelihood confidence intervals for the difference of two areas under the ROC curves with right censored data, for different censored rates, sample sizes, and nominal levels.

Let the diseased population $X = (X_1, X_2)$ and the non-diseased population $Y = (Y_1, Y_2)$ distributed as the same settings in Section 3.3. For a certain censoring rate and a certain sample size, 1000 independent random samples of data $\{(x_{ki}, \delta_{x_{ki}}), i = 1, ..., m; (y_{kj}, \delta_{y_{kj}}), j = 1, ..., n, k = 1, 2\}$ are generated. The nominal levels of the confidence intervals are $1-\alpha = 95\%$ and $1-\alpha = 90\%$.

From Tables 5.1 - 5.4, we have the following observations:

1. For each censoring rate, sample size and nominal levels, the coverage probability is close to the nominal level, and the average lengths of the empirical likelihood based confidence intervals are short;

2. In almost all the scenarios, as the censoring rates decrease or the sample sizes increase, the coverage probabilities get closer to the nominal level, and the average lengths of the intervals decreases respectively. This is reasonable since either smaller censoring rates or bigger sample sizes provide more information of the data under study.

5.4 Real Application

In this section, the proposed EL method is illustrated by the same PBC data set as in Chapter 3. We construct 90% and 95% confidence intervals for the difference of 2 AUC's. Similarly to Wang et al. (2009), we use the estimates of AUC's instead of the true values of AUC's, and the difference of estimates is employed instead of the true value of the difference. We utilize the bootstrap method to improve the accuracy, and $B = 400$ in our data analysis. The 95% EL confidence interval is $(0.062, 0.377)$, and the 90% EL confidence interval is $(0.088, 0.353)$. Both lower bounds are larger than zero, which implies that the AUC for the group with hepatomegaly is larger than that for the group without hepatomegaly. This conclusion is consistent with the result we have made in section 3.4.

Table 5.1: Empirical likelihood confidence intervals for the difference of two AUCs with right censoring data at the nominal level of $1-\alpha = 95\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(\lambda_x)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(\lambda_y)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.2$.

(m_1, m_2)	(n_1, n_2)	(λ_x, λ_y)	EL	
			CP(%)	AL
(50, 50)	(50, 50)	(8, 2)	94.9	0.296
		(6, 4)	94.8	0.326
		(4, 6)	94.2	0.325
		(2, 8)	94.3	0.287
(100, 100)	(100, 100)	(8, 2)	93.9	0.207
		(6, 4)	94.4	0.229
		(4, 6)	94.8	0.228
		(2, 8)	94.8	0.203
(150, 150)	(150, 150)	(8, 2)	95.1	0.167
		(6, 4)	95.4	0.186
		(4, 6)	95.4	0.186
		(2, 8)	94.7	0.165

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 5.2: Empirical likelihood confidence intervals for the difference of two AUCs with right censoring data at the nominal level of $1-\alpha = 90\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(\lambda_x)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(\lambda_y)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.2$.

(m_1, m_2)	(n_1, n_2)	(λ_x, λ_y)	EL	
			CP(%)	AL
(50, 50)	(50, 50)	(8, 2)	90.9	0.248
		(6, 4)	90.1	0.272
		(4, 6)	90.3	0.271
		(2, 8)	90.0	0.240
(100, 100)	(100, 100)	(8, 2)	89.9	0.173
		(6, 4)	89.3	0.191
		(4, 6)	88.6	0.191
		(2, 8)	87.9	0.169
(150, 150)	(150, 150)	(8, 2)	90.4	0.140
		(6, 4)	90.9	0.156
		(4, 6)	91.0	0.155
		(2, 8)	90.5	0.138

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 5.3: Empirical likelihood confidence intervals for the difference of two AUCs with right censoring data at the nominal level of $1-\alpha = 95\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(\lambda_x)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(\lambda_y)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.1$.

(m_1, m_2)	(n_1, n_2)	(λ_x, λ_y)	EL	
			CP(%)	AL
(50, 50)	(50, 50)	(8, 2)	94.0	0.275
		(6, 4)	94.6	0.307
		(4, 6)	94.2	0.307
		(2, 8)	94.4	0.274
(100, 100)	(100, 100)	(8, 2)	94.7	0.195
		(6, 4)	94.0	0.218
		(4, 6)	94.4	0.218
		(2, 8)	94.3	0.195
(150, 150)	(150, 150)	(8, 2)	94.4	0.159
		(6, 4)	94.9	0.178
		(4, 6)	94.9	0.179
		(2, 8)	95.5	0.160

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

Table 5.4: Empirical likelihood confidence intervals for the difference of two AUCs with right censoring data at the nominal level of $1-\alpha = 90\%$, $X_1 \sim exp(4)$, $X_2 \sim exp(\lambda_x)$, $Y_1 \sim exp(2)$, $Y_2 \sim exp(\lambda_y)$, and the censoring rates $c_{X_1} = c_{X_2} = c_{Y_1} = c_{Y_2} = 0.1$.

(m_1, m_2)	(n_1, n_2)	(λ_x, λ_y)	EL	
			CP(%)	AL
(50, 50)	(50, 50)	(8, 2)	88.2	0.230
		(6, 4)	89.4	0.257
		(4, 6)	89.5	0.257
		(2, 8)	89.4	0.230
(100, 100)	(100, 100)	(8, 2)	89.4	0.163
		(6, 4)	88.6	0.183
		(4, 6)	88.7	0.183
		(2, 8)	89.7	0.163
(150, 150)	(150, 150)	(8, 2)	91.0	0.134
		(6, 4)	90.7	0.150
		(4, 6)	90.8	0.149
		(2, 8)	89.6	0.134

NOTE:
EL: empirical likelihood,
CP(%): coverage probability,
AL: average length of a confidence interval.

CHAPTER 6

JACKKNIFE EMPIRICAL LIKELIHOOD FOR THE DIFFERENCE OF 2 VOLUMES UNDER ROC SURFACES

6.1 Background

A multi-category classification technique is necessary if the subjects are supposed to be assigned to more than two groups simultaneously. Mossman (1999) evaluated a three-category classification treatment using the volume under the ROC surface (VUS). It is proposed as an analogous measure to the AUC, extending an ROC curve to an ROC surface in a three dimensional case. Also, the VUS provides a scalar measure as the AUC does. Tian et al. (2011) showed that the difference of two correlated VUS's is an efficient summary for the comparison of diagnostic accuracy with three ordinal diagnostic groups using parametric methods. Here the 'difference' between two VUS's implies the amount by which one of them is subtracted by the other.

Since a VUS identifies a three-category data as in Wan (2012), where it has one more category than an ROC curve or an AUC, the estimating equations for the difference of two VUS's are much more complex than the difference of two ROC curves or that of two AUC's, even for complete data. Thus, it is rather difficult to construct a confidence interval of such difference until Jing et al. (2009) introduced the jackknife empirical likelihood (JEL) method making such kind of problems tractable. JEL employs a U-statistic to avoid the nuisance parameters in the estimating equations, it therefore provides a reliable confidence interval by solving a simpler estimating equation of a pseudo mean, which is based on U-statistics (Korolyuk and Borovskikh (1994)). The orginal JEL considers univariate problems. Pan et al. (2013) made nonparametric statistical inference for the VUS's using JEL.

In this paper, we proposed a novel U-statistic for the JEL method to deal with the difference of trivariate problems. Our results show that JEL confidence intervals outperform the normal approximation (NA) method for the difference of two correlated VUS's, as Owen (1988, 1990)'s empirical likelihood (EL) method is too complicated to be employed.

The rest of this chapter is organized as follows. In Section 6.2, the JEL method is employed to construct the confidence intervals for the difference of two VUS's. We prove that the limiting distribution of the empirical log-likelihood ratio statistic follows a χ^2-distribution. In Section 6.3, we present the results of intensive simulation studies on the JEL confidence intervals, which have better performance than those based on the NA method in terms of coverage probability. In Section 6.4, the proposed method is illustrated by an Alzheimer's Disease (AD) data. All the proofs are provided in Appendix E.

6.2 Main Results

6.2.1 The Difference of Two VUS's

Let $(X_1^T, X_2^T, ..., X_{n_1}^T)$, $(Y_1^T, Y_2^T, ..., Y_{n_2}^T)$ and $(Z_1^T, Z_2^T, ..., Z_{n_3}^T)$ represent i.i.d. samples of three independent populations, where $X_i = (X_{1i}, X_{2i})^T$, $i = 1, 2, ..., n_1$, $Y_j = (Y_{1j}, Y_{2j})^T$, $j = 1, 2, ..., n_2$, and $Z_k = (Z_{1k}, Z_{2k})^T$, $k = 1, 2, ..., n_3$. We define the VUS with respect to the first component as $P(X_{11} < Y_{11} < Z_{11})$, and the VUS with respect to the second component as $P(X_{21} < Y_{21} < Z_{21})$, respectively. Therefore the difference of two VUS's can be defined as

$$\begin{aligned}\theta =& P(X_{11} < Y_{11} < Z_{11}) - P(X_{21} < Y_{21} < Z_{21}) \\ =& E(I(X_{11} < Y_{11} < Z_{11})) - E(I(X_{21} < Y_{21} < Z_{21})) \\ =& E(I(X_{11} < Y_{11} < Z_{11}) - I(X_{21} < Y_{21} < Z_{21})),\end{aligned}$$

which can be estimated by

$$\hat{\theta} = \frac{1}{n_1 n_2 n_3} \sum_{\substack{i=1,...,n_1, \\ j=1,...,n_2, \\ k=1,...,n_3}} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})].$$

6.2.2 U-statistics

A U-statistic of degree $(1, 1, 1)$ with a kernel $h(x; y; z)$ is defined as

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} h(X_i; Y_j; Z_k),$$

which is a consistent and unbiased estimator of the parameter $\theta = Eh(X_i; Y_j; Z_k)$. In particular, if

$$h(X_i; Y_j; Z_k) = I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k}),$$

then $\theta = P(X_{11} < Y_{11} < Z_{11}) - P(X_{21} < Y_{21} < Z_{21})$. Therefore we define the U-statistic for inference on θ as

$$U_n = \frac{1}{n_1 n_2 n_3} \sum_{i=1}^{n1} \sum_{j=1}^{n2} \sum_{k=1}^{n3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})].$$

In addition, for $i = 1, 2, ..., n_1$, $j = 1, 2, ..., n_2$, and $k = 1, 2, ..., n_3$, we denote

(1) the original statistics for all observations as $U^0_{n_1,n_2,n_3} = U_n$;

(2) the statistic after removing $X_{i'}$ as

$$U^{-i',0,0}_{n_1-1,n_2,n_3} = ((n_1-1)n_2n_3)^{-1} \sum_{\substack{i=1,\\ i\neq i'}}^{n_1} \sum_{j=1}^{n_2} \sum_{k=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];$$

(3) the statistic after removing $Y_{j'}$ as

$$U^{0,-j',0}_{n_1,n_2-1,n_3} = (n_1(n_2-1)n_3)^{-1} \sum_{i=1}^{n_1} \sum_{\substack{j=1,\\ j\neq j'}}^{n_2} \sum_{k=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];$$

(4) the statistic after removing $Z_{k'}$ as

$$U^{0,0,-k'}_{n_1,n_2,n_3-1} = (n_1n_2(n_3-1))^{-1} \sum_{i=1}^{n_1} \sum_{j=1}^{n2} \sum_{\substack{k=1,\\ k\neq k'}}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})].$$

6.2.3 JEL for the Difference of Two VUS's

Hence we define the jackknife pseudo-values by

$$V_{i,0,0} = n_1 U^0_{n_1,n_2,n_3} - (n_1 - 1) U^{-i,0,0}_{n_1-1,n_2,n_3};$$

$$V_{0,j,0} = n_2 U^0_{n_1,n_2,n_3} - (n_2 - 1) U^{0,-j,0}_{n_1,n_2-1,n_3};$$

$$V_{0,0,k} = n_3 U^0_{n_1,n_2,n_3} - (n_3 - 1) U^{0,0,-k}_{n_1,n_2,n_3-1}.$$

With some simple algebra,

$$V_{i,0,0} = \frac{1}{n_2 n_3} \sum_{j_1=1}^{n_2} \sum_{k_1=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];$$

$$V_{0,j,0} = \frac{1}{n_1 n_3} \sum_{i_1=1}^{n_1} \sum_{k_1=1}^{n_3} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];$$

$$V_{0,0,k} = \frac{1}{n_1 n_2} \sum_{i_1=1}^{n_1} \sum_{j_1=1}^{n_2} [I(X_{1i} < Y_{1j} < Z_{1k}) - I(X_{2i} < Y_{2j} < Z_{2k})];$$

and

$$\bar{V}_{\cdot,0,0} = \bar{V}_{0,\cdot,0} = \bar{V}_{0,0,\cdot} = U_n,$$

where $\bar{V}_{\cdot,0,0}$, $\bar{V}_{0,\cdot,0}$ and $\bar{V}_{0,0,\cdot}$ are the averages of $V_{i,0,0}$, $V_{0,j,0}$ and $V_{0,0,k}$, respectively.

We will also need the following notations for the inferences in the paper:

$$g_{1,0,0}(x) = [P(x_{11} < Y_{11} < Z_{11}) - P(x_{21} < Y_{21} < Z_{21})] - \theta,\ \sigma^2_{1,0,0} = Var(g_{1,0,0}(X_1));$$

$$g_{0,1,0}(y) = [P(X_{11} < y_{11} < Z_{11}) - P(X_{21} < y_{21} < Z_{21})] - \theta,\ \sigma^2_{0,1,0} = Var(g_{0,1,0}(Y_1));$$

$$g_{0,0,1}(z) = [P(X_{11} < Y_{11} < z_{11}) - P(X_{21} < Y_{21} < z_{21})] - \theta,\ \sigma^2_{0,0,1} = Var(g_{0,0,1}(Z_1));$$

where $x = (x_1, x_2)^T$, $y = (y_1, y_2)^T$, and $z = (z_1, z_2)^T$.

Denote $(T_1, T_2, ..., T_n) = (T_1, T_2, ..., T_{n_1}, T_{n_1+1}, T_{n_1+2}, ..., T_{n_1+n_2}, T_{n_1+n_2+1}, ..., T_{n_1+n_2+n_3})$

$= (X_1^T, X_2^T, ..., X_{n_1}^T, Y_1^T, Y_2^T, ..., Y_{n_2}^T, Z_1^T, Z_2^T, ..., Z_{n_3}^T)$, where $n = n_1 + n_2 + n_3$. A one-sample U-statistic of degree 3 is defined as

$$W_n = U_n(T_1, T_2, ..., T_n) = \binom{n}{3}^{-1} \sum_{1 \le i < j < k \le n} h(T_i, T_j, T_k),$$

where the kernel function

$$h(T_i, T_j, T_k) = \frac{\binom{n}{3}}{n_1 n_2 n_3}[I(X_{1i} < Y_{1,j-n_1} < Z_{1,k-n_1-n_2}) - I(X_{2i} < Y_{2,j-n_1} < Z_{2,k-n_1-n_2})] \quad (6.1)$$

for $i = 1, 2, ..., n_1,\ j = n_1 + 1, n_1 + 2, ..., n_1 + n_2,\ k = n_1 + n_2 + 1, n_1 + n_2 + 2, ..., n$, and $1 \le i \le n_1 < j \le n_1 + n_2 < k \le n$, and $h(T_i, T_j, T_k) = 0$ otherwise. We can use the U-statistic as an unbiased estimator of the parameter θ. Note that $\theta = Eh(T_i, T_j, T_k)$, and $W_n = U_n$. Define the U-statistics with T_l deleted as follows:

$$\begin{aligned}
&W_{n-1}^{(-l)} \\
=&U_{n-1}(T_1, T_2, ..., T_{l-1}, T_{l+1}, ..., T_n) \\
=&\binom{n-1}{3}^{-1} \sum_{n-1,3}^{(-l)} h(T_i, T_j, T_k) \\
=&\binom{n-1}{3}^{-1} [\sum_{i<j<k} h(T_i, T_j, T_k) - \sum_{j<k} h(T_l, T_j, T_k) - \sum_{i<k} h(T_i, T_l, T_k) - \sum_{i<j} h(T_i, T_j, T_l)] \\
=&\binom{n-1}{3}^{-1} \Big[\binom{n}{3} W_n - \sum_{j<k} h(T_l, T_j, T_k) - \sum_{i<k} h(T_i, T_l, T_k) - \sum_{i<j} h(T_i, T_j, T_l)\Big],
\end{aligned}$$

where $1 \le l \le n$, denote the removal of T_l as $(-l)$.

Hence we define the jackknife pseudo-values by

$$
\begin{aligned}
\hat{V}_l =& nW_n - (n-1)W_{n-1}^{(-l)} \\
=& nW_n - (n-1)\binom{n-1}{3}^{-1}\binom{n}{3}W_n \\
& + (n-1)\binom{n-1}{3}^{-1}[\sum_{l<j<k} h(T_l,T_j,T_k) + \sum_{i<l<k} h(T_i,T_l,T_k) + \sum_{i<j<l} h(T_i,T_j,T_l)] \\
=& -\frac{2n}{n-3}U_n + \frac{6}{(n-2)(n-3)}[\sum_{l<j<k} h(T_l,T_j,T_k) + \sum_{i<l<k} h(T_i,T_l,T_k) + \sum_{i<j<l} h(T_i,T_j,T_l)].
\end{aligned}
$$

Now plugging in equation (6.1), we have

$$
\begin{aligned}
& \hat{V}_l \\
=& -\frac{2n}{n-3}U_n + \frac{6}{(n-2)(n-3)}\frac{n(n-1)(n-2)}{6n_1n_2n_3} \\
& \{\sum_{j<k}[I(X_{1l} < Y_{1,j-n_1} < Z_{1,k-n_1-n_2}) - I(X_{2l} < Y_{2,j-n_1} < Z_{2,k-n_1-n_2})] \\
& I(1 \le l \le n_1 < j \le n_1+n_2 < k \le n) \\
& + \sum_{i<k}[I(X_{1i} < Y_{1l} < Z_{1,k-n_1-n_2}) - I(X_{2i} < Y_{2l} < Z_{2,k-n_1-n_2})] \\
& I(1 \le i \le n_1 < l \le n_1+n_2 < k \le n) \\
& + \sum_{i<j}[I(X_{1i} < Y_{1,j-n_1} < Z_{1,l}) - I(X_{2i} < Y_{2,j-n_1} < Z_{2l})] \\
& I(1 \le i \le n_1 < j \le n_1+n_2 < l \le n)\} \\
=& -\frac{2n}{n-3}U_n + \frac{n(n-1)}{(n-3)}\frac{1}{n_1n_2n_3} \\
& \{\sum_{j=n_1+1}^{n_1+n_2}\sum_{k=n_1+n_2+1}^{n}[I(X_{1l} < Y_{1,j-n_1} < Z_{1,k-n_1-n_2}) - I(X_{2l} < Y_{2,j-n_1} < Z_{2,k-n_1-n_2})]I(1 \le l \le n_1) \\
& + \sum_{i=1}^{n_1}\sum_{k=n_1+n_2+1}^{n}[I(X_{1i} < Y_{1l} < Z_{1,k-n_1-n_2}) - I(X_{2i} < Y_{2l} < Z_{2,k-n_1-n_2})]I(n_1 < l \le n_1+n_2) \\
& + \sum_{i=1}^{n_1}\sum_{j=n_1+1}^{n_1+n_2}[I(X_{1i} < Y_{1,j-n_1} < Z_{1,l}) - I(X_{2i} < Y_{2,j-n_1} < Z_{2l})]I(n_1+n_2 < l \le n)\}.
\end{aligned}
$$

Therefore,

$$E(\hat{V}_l) = -\frac{2n}{n-3}\theta + \frac{n(n-1)}{(n-3)}\Big[\frac{\theta}{n_1}I(1 \le l \le n_1) + \frac{\theta}{n_2}I(n_1 < l \le n_1+n_2) + \frac{\theta}{n_3}I(n_1+n_2 < l \le n)\Big].$$

Let $p = (p_1, p_2, ..., p_n)$ be a probability vector, i.e., $\sum_{i=1}^{n} p_i = 1$ and $p_i \ge 0$ for $1 \le i \le n$. By employing the idea of Jing et al. (2009), the jackknife empirical likelihood ratio function for θ is

$$R(\theta) = \sup_{p_1,\ldots,p_n,} \{\prod_{i=1}^{n}(np_i) | p_i > 0, \sum_{i=1}^{n} p_i = 1, \sum_{i=1}^{n} p_i \hat{V}_i - \sum_{i=1}^{n} p_i E\hat{V}_i = 0\}.$$

Using Lagrange multiplier method, we have $\log R(\theta) = -\sum_{l=1}^{n} \log(1 + \gamma(\hat{V}_l - E\hat{V}_l))$, where γ is the solution to the equation

$$\frac{1}{n}\sum_{l=1}^{n} \frac{\hat{V}_l - E\hat{V}_l}{1 + \gamma(\hat{V}_l - E\hat{V}_l)} = 0. \tag{6.2}$$

The Wilk's theorem holds for θ.

Theorem 6.1. *Assume that*

(a) $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$, $\sigma^2_{0,0,1} > 0$, *and*

(b) $\frac{n}{n_t} \to c_t < \infty$, *where* $t = 1, 2, 3$ *and* c_t*'s are finite constants.*

The empirical log-likelihood ratio statistic at the true value θ_0

$$l(\theta_0) = -2\log R(\theta_0) \xrightarrow{d} \chi^2_1,$$

as $\min(n_1, n_2, n_3) \to \infty$, *where* χ^2_1 *is a standard* χ^2*-distribution with degree of freedom 1.*

Thus, the asymptotic $100(1-\alpha)\%$ jackknife empirical likelihood confidence interval for θ is given by

$$\{\theta : l(\theta) \le \chi^2_1(\alpha)\},$$

where $\chi^2_1(\alpha)$ is the upper α-quantile of χ^2_1.

6.3 Simulation Study

In this section, we conduct extensive simulation studies to investigate the finite sample performance of the proposed jackknife empirical likelihood method for the difference of two VUS's with different data sets. For comparison purpose, we also construct the confidence intervals based on the normal approximation method. We present the conclusion for the normal approximation method at Lemma E.1 in the Appendix. Based on Lemma E.1, the $100(1-\alpha)\%$ confidence intervals based on the normal approximation method can be constructed as

$$I = \left\{\theta : |U_n - \theta| \leq Z_{\alpha/2}\hat{\sigma}\right\},$$

where $Z_{\alpha/2}$ is the upper $\alpha/2$ critical value for the standard normal distribution, and $\hat{\sigma}$ is defined in Appendix A. In this paper, we evaluate the proposed methods in terms of average length and coverage probability of confidence intervals.

For Table 6.1 and Table 6.2, the data follow the Marshall-Olkin bivariate exponential distribution (MOBVE), as in Marshall and Olkin (1967) and Balakrishnan (1996). $MOBVE(\lambda_1, \lambda_2, \lambda_3)$ has a CDF

$$F(w_1, w_2) = 1 - exp[-\lambda_1 w_1 - \lambda_2 w_2 - \lambda_3 max\{w_1, w_2\}],$$

where $w_1, w_2 > 0$, $\lambda_t \geq 0$ and at least one λ_t is positive, $t = 1, 2, 3$. The marginal distributions of (W_1, W_2) are exponential with expectations $(\lambda_1 + \lambda_3)$ and $(\lambda_2 + \lambda_3)$, respectively. Their correlation c is $\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$. In this simulation study, the first population $X = (X_1, X_2) = (\rho_x X_1^*, X_2^*)$, where $(X_1^*, X_2^*) \sim MOBVE(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3})$, and $\rho_x = 3$. The second population $Y = (Y_1, Y_2) = (\rho_y Y_1^*, Y_2^*)$, where $(Y_1^*, Y_2^*) \sim MOBVE(\lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3})$, and $\rho_y = 2$. The third population $Z = (Z_1, Z_2) = (\rho_z Z_1^*, Z_2^*)$, where $(Z_1^*, Z_2^*) \sim MOBVE(\lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$, and $\rho_z = 1$. The $\lambda_{x_t}, \lambda_{y_t}, \lambda_{z_t}$'s differ for various correlations, where the correlations c_1, c_2 and c_3 are chosen as 0, 0.25, 0.5, 0.75, and 0.9. Additionally, we guarantee the marginal distributions $X_1^* \sim exp(1)$, $X_2^* \sim exp(1)$, $Y_1^* \sim exp(2)$, $Y_2^* \sim exp(2)$, $Z_1^* \sim exp(3)$, and

$Z_2^* \sim exp(3)$.

In Table 6.3 and Table 6.4, the data are generated from the bivariate normal distributions. The distributions are: $(X_1, X_2) \sim N(\mu_x, \Sigma_x)$, $(Y_1, Y_2) \sim N(\mu_y, \Sigma_y)$, $(Z_1, Z_2) \sim N(\mu_z, \Sigma_z)$, where $\mu_x = (5, 3)$, $\mu_y = (4, 2)$, and $\mu_z = (4, 2)$, and the covariance matrices are

$$\Sigma_x = \Sigma_y = \Sigma_z = \begin{pmatrix} 1 & c \\ c & 1 \end{pmatrix}$$

as the correlation c varies.

The sample sizes for x, y and z of $(n_{x_1}, n_{x_2}, n_{y_1}, n_{y_2}, n_{z_1}, n_{z_2})$ are $(10, 10, 10, 10, 10, 10)$, $(20, 20, 25, 25, 30, 30)$, $(30, 30, 30, 30, 30, 30)$, $(60, 60, 60, 60, 60, 60)$, $(80, 80, 80, 80, 80, 80)$ and $(100, 100, 100, 100, 100, 100)$. The nominal levels of the confidence intervals are $1 - \alpha = 95\%$ and $1 - \alpha = 90\%$. For a certain correlation and certain sample size, 1000 iterations are repeated.

From Tables 6.1 - 6.4 we make the following conclusions:

1. For different correlations, sample sizes and parameters of the distributions, the coverage probabilities of the confidence intervals based on the JEL methods and those based on the NA methods are close to the nominal levels;

2. In almost all the scenarios, as the sample sizes increase, the coverage probabilities of the confidence intervals for the two methods get closer to the nominal level, and the average lengths of the intervals decrease respectively. This is reasonable since larger sample sizes provide more information of the data under study;

3. For the same sample sizes, as the correlations increase, the coverage probabilities of the confidence intervals for the two methods are closer to the nominal level, and the average lengths of the intervals decrease respectively. Jackknife empirical likelihood based confidence intervals outperform the normal approximation method for various sample sizes and correlation coefficients.

6.4 Real Application

In this section, the proposed method for the confidence intervals of the difference of two VUS's is illustrated via a data set of the diagnosis for early stage Alzheimer's Disease (AD) from the Alzheimer's Disease Research Center (ADRC) at Washington University (See Xiong et al. (2006)). The severity of dementia of Alzheimer type could be staged by the clinical dementia rating (CDR), a score based on several clinical evaluations and neuropsychometric measurements. The purpose of the study is to investigate the early stage Alzheimer's Disease. Thus we concentrate on the following three diagnostic groups: non-demented group (CDR 0), very mildly demented group (CDR 0.5), and mildly demented group (CDR 1). The data set includes 14 neuropsychometric markers from 118 cases aged 75 falling into the three diagnostic categories mentioned above. Out of the 14 measures, we compare the diagnostic accuracies between the scores from two neuropsychometric tests. One of them is a measure of semantic memory, named as the Information subset of the Wechsler Adult Intelligence Scale (WAIS), see Wechsler (1955). The other is an untimed visuospatial measure called Visual Retention Test (Form D, copy), as in Storandt and Hill (1989).

By deleting the individuals with results of missing values, we have 22 patients from mildly demented group (CDR 1), 44 patients from very mildly demented group (CDR 0.5), and 45 participants from non-demented group (CDR 0).

For CDR 1 group, the sample mean is $(-2.125, -1.769)$, and the sample covariance matrix is

$$\begin{pmatrix} 1.298 & 0.786 \\ 0.786 & 5.751 \end{pmatrix}.$$

The correlation of the two attributes is 0.288.

For CDR 0.5 group, the sample mean is $(-0.607, -0.551)$, and the sample covariance matrix is

$$\begin{pmatrix} 1.167 & 1.302 \\ 1.302 & 3.476 \end{pmatrix}.$$

The correlation of the two attributes is 0.647.

For CDR 0 group, the sample mean is $(0.631, 0.202)$, and the sample covariance matrix is

$$\begin{pmatrix} 0.712 & 0.164 \\ 0.164 & 0.445 \end{pmatrix}.$$

The correlation of the two attributes is 0.292.

The interval estimate of the difference of the two VUS's based on the JEL method is $(0.350, 0.634)$ at $\alpha = 90\%$, and $(0.324, 0.662)$ at $\alpha = 95\%$. The NA confidence interval is $(0.375, 0.604)$ at $\alpha = 90\%$, and $(0.353, 0.627)$ at $\alpha = 95\%$. Therefore, we can conclude that the Information subset of the WAIS possesses a stronger discrimination power than that of Visual Retention Test (Form D, copy).

Table 6.1: Jackknife empirical likelihood confidence intervals for the difference of volume under ROC surfaces (VUS) at the nominal level of $1-\alpha = 95\%$. The distributions are $(X_1^*, X_2^*) \sim MOBVE(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3})$, $(Y_1^*, Y_2^*) \sim MOBVE(\lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3})$, $(Z_1^*, Z_2^*) \sim MOBVE(\lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$. The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1$, $n_{y_1} = n_{y_2} = n_2$, $n_{z_1} = n_{z_2} = n_3$.

c	$(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3}; \lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3}; \lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$	(n_1, n_2, n_3)	JEL		NA	
			CP(%)	AL	CP(%)	AL
0	(1, 1, 0; 2, 2, 0; 3, 3, 0)	(10, 10, 10)	90.8	.203	90.0	.190
		(20, 25, 30)	94.1	.126	92.4	.120
		(30, 30, 30)	94.8	.108	93.6	.104
		(60, 60, 60)	94.6	.074	94.5	.072
		(80, 80, 80)	94.5	.064	94.7	.063
		(100,100,100)	95.5	.057	95.3	.056
0.25	$(\frac{3}{5}, \frac{3}{5}, \frac{2}{5}; \frac{6}{5}, \frac{6}{5}, \frac{4}{5}; \frac{9}{5}, \frac{9}{5}, \frac{6}{5})$	(10, 10, 10)	91.1	.188	89.6	.176
		(20, 25, 30)	94.4	.112	93.5	.106
		(30, 30, 30)	94.9	.096	93.7	.093
		(60, 60, 60)	95.9	.066	95.2	.064
		(80, 80, 80)	94.2	.056	93.5	.055
		(100,100,100)	94.2	.050	93.6	.049
0.5	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}; \frac{2}{3}, \frac{2}{3}, \frac{4}{3}; 1, 1, 2)$	(10, 10, 10)	89.6	.167	86.9	.158
		(20, 25, 30)	94.4	.100	93.0	.094
		(30, 30, 30)	94.2	.087	93.0	.083
		(60, 60, 60)	92.8	.058	90.8	.056
		(80, 80, 80)	95.3	.049	94.6	.049
		(100,100,100)	95.7	.044	95.1	.044
0.75	$(\frac{1}{7}, \frac{1}{7}, \frac{6}{7}; \frac{2}{7}, \frac{2}{7}, \frac{12}{7}; \frac{3}{7}, \frac{3}{7}, \frac{18}{7})$	(10, 10, 10)	90.1	.151	87.1	.144
		(20, 25, 30)	95.1	.088	92.4	.084
		(30, 30, 30)	95.9	.075	93.1	.073
		(60, 60, 60)	94.8	.051	93.6	.050
		(80, 80, 80)	95.9	.044	94.7	.043
		(100,100,100)	93.5	.039	92.2	.038
0.9	$(\frac{1}{19}, \frac{1}{19}, \frac{18}{19}; \frac{2}{19}, \frac{2}{19}, \frac{36}{19}; \frac{3}{19}, \frac{3}{19}, \frac{54}{19})$	(10, 10, 10)	93.5	.143	90.1	.136
		(20, 25, 30)	94.6	.080	92.3	.078
		(30, 30, 30)	94.7	.069	93.7	.067
		(60, 60, 60)	95.4	.046	94.9	.046
		(80, 80, 80)	94.9	.040	94.6	.039
		(100,100,100)	95.4	.035	94.6	.035

NOTE:
JEL: Jackknife Empirical Likelihood,
NA: Normal Approximation,
CP(%): Coverage Probability,
AL: Average Length.

Table 6.2: Jackknife empirical likelihood confidence intervals for the difference of volume under ROC surfaces (VUS) at the nominal level of $1-\alpha = 90\%$. The distributions are $(X_1^*, X_2^*) \sim MOBVE(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3})$, $(Y_1^*, Y_2^*) \sim MOBVE(\lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3})$, $(Z_1^*, Z_2^*) \sim MOBVE(\lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$. The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1$, $n_{y_1} = n_{y_2} = n_2$, $n_{z_1} = n_{z_2} = n_3$.

c	$(\lambda_{x_1}, \lambda_{x_2}, \lambda_{x_3}; \lambda_{y_1}, \lambda_{y_2}, \lambda_{y_3}; \lambda_{z_1}, \lambda_{z_2}, \lambda_{z_3})$	(n_1, n_2, n_3)	JEL		NA	
			CP(%)	AL	CP(%)	AL
0	(1, 1, 0; 2, 2, 0; 3, 3, 0)	(10, 10, 10)	87.2	.168	85.0	.190
		(20, 25, 30)	88.7	.105	87.8	.101
		(30, 30, 30)	89.8	.090	88.3	.087
		(60, 60, 60)	89.7	.062	88.9	.061
		(80, 80, 80)	90.2	.053	89.4	.053
		(100,100,100)	88.8	.047	88.6	.047
0.25	$(\frac{3}{5}, \frac{3}{5}, \frac{2}{5}; \frac{6}{5}, \frac{6}{5}, \frac{4}{5}; \frac{9}{5}, \frac{9}{5}, \frac{6}{5})$	(10, 10, 10)	86.9	.156	85.8	.148
		(20, 25, 30)	90.6	.093	88.9	.089
		(30, 30, 30)	90.4	.080	89.5	.078
		(60, 60, 60)	90.5	.055	90.3	.054
		(80, 80, 80)	90.0	.047	89.8	.046
		(100,100,100)	90.3	.042	88.7	.041
0.5	$(\frac{1}{3}, \frac{1}{3}, \frac{2}{3}; \frac{2}{3}, \frac{2}{3}, \frac{4}{3}; 1, 1, 2)$	(10, 10, 10)	85.7	.139	83.7	.133
		(20, 25, 30)	90.0	.083	87.5	.080
		(30, 30, 30)	90.5	.072	89.5	.071
		(60, 60, 60)	87.3	.048	86.4	.047
		(80, 80, 80)	91.2	.041	90.8	.041
		(100,100,100)	90.0	.037	89.9	.037
0.75	$(\frac{1}{7}, \frac{1}{7}, \frac{6}{7}; \frac{2}{7}, \frac{2}{7}, \frac{12}{7}; \frac{3}{7}, \frac{3}{7}, \frac{18}{7})$	(10, 10, 10)	86.3	.126	83.3	.121
		(20, 25, 30)	90.4	.073	89.1	.070
		(30, 30, 30)	91.4	.062	89.8	.061
		(60, 60, 60)	88.5	.042	87.6	.042
		(80, 80, 80)	91.0	.037	89.2	.036
		(100,100,100)	87.7	.032	87.3	.032
0.9	$(\frac{1}{19}, \frac{1}{19}, \frac{18}{19}; \frac{2}{19}, \frac{2}{19}, \frac{36}{19}; \frac{3}{19}, \frac{3}{19}, \frac{54}{19})$	(10, 10, 10)	90.0	.119	87.1	.114
		(20, 25, 30)	91.0	.067	88.7	.065
		(30, 30, 30)	91.3	.057	89.1	.056
		(60, 60, 60)	90.1	.039	89.6	.038
		(80, 80, 80)	91.9	.033	91.0	.033
		(100,100,100)	90.2	.029	90.0	.029

NOTE:
JEL: Jackknife Empirical Likelihood,
NA: Normal Approximation,
CP(%): Coverage Probability,
AL: Average Length.

Table 6.3: Jackknife empirical likelihood confidence intervals for the difference of volume under ROC surfaces (VUS) at the nominal level of $1-\alpha = 95\%$. The marginal distributions are $X_1 \sim N(\mu_{x_1}, 1)$, $X_2 \sim N(\mu_{x_2}, 1)$, $Y_1 \sim N(\mu_{y_1}, 1)$, $Y_2 \sim N(\mu_{y_2}, 1)$, $Z_1 \sim N(\mu_{z_1}, 1)$, $Z_2 \sim N(\mu_{z_2}, 1)$. The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1$, $n_{y_1} = n_{y_2} = n_2$, $n_{z_1} = n_{z_2} = n_3$.

c	$(\mu_{x_1}, \mu_{x_2}, \mu_{y_1}, \mu_{y_2}, \mu_{z_1}, \mu_{z_2})$	(n_1, n_2, n_3)	JEL		NA	
			CP(%)	AL	CP(%)	AL
0	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	90.1	.182	89.2	.171
		(20, 25, 30)	93.6	.112	92.7	.106
		(30, 30, 30)	94.8	.097	93.3	.093
		(60, 60, 60)	95.7	.067	94.0	.065
		(80, 80, 80)	95.4	.057	94.8	.056
		(100,100,100)	94.8	.051	94.5	.050
0.25	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	91.3	.176	90.0	.165
		(20, 25, 30)	94.1	.107	93.1	.101
		(30, 30, 30)	93.9	.092	93.5	.088
		(60, 60, 60)	94.6	.063	93.7	.062
		(80, 80, 80)	95.1	.054	94.3	.053
		(100,100,100)	94.6	.048	92.3	.047
0.5	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	89.2	.163	88.0	.153
		(20, 25, 30)	93.6	.101	92.1	.096
		(30, 30, 30)	93.9	.084	91.4	.081
		(60, 60, 60)	94.7	.058	93.9	.056
		(80, 80, 80)	95.1	.050	94.6	.049
		(100,100,100)	95.2	.044	93.8	.043
0.75	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	89.6	.148	87.6	.140
		(20, 25, 30)	93.9	.088	91.4	.084
		(30, 30, 30)	94.7	.075	91.5	.072
		(60, 60, 60)	94.9	.051	93.3	.050
		(80, 80, 80)	95.3	.043	94.0	.042
		(100,100,100)	95.5	.039	95.0	.038
0.9	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	91.3	.134	88.4	.127
		(20, 25, 30)	93.8	.079	91.3	.075
		(30, 30, 30)	95.6	.067	92.3	.065
		(60, 60, 60)	95.6	.046	94.8	.045
		(80, 80, 80)	95.6	.039	95.1	.038
		(100,100,100)	95.1	.035	94.7	.034

NOTE:
JEL: Jackknife Empirical Likelihood,
NA: Normal Approximation,
CP(%): Coverage Probability,
AL: Average Length.

Table 6.4: Jackknife empirical likelihood confidence intervals for the difference of volume under ROC surfaces (VUS) at the nominal level of $1-\alpha = 90\%$. The marginal distributions are $X_1 \sim N(\mu_{x_1}, 1)$, $X_2 \sim N(\mu_{x_2}, 1)$, $Y_1 \sim N(\mu_{y_1}, 1)$, $Y_2 \sim N(\mu_{y_2}, 1)$, $Z_1 \sim N(\mu_{z_1}, 1)$, $Z_2 \sim N(\mu_{z_2}, 1)$. The correlations $c_1 = c_2 = c_3 = c$, and sample sizes $n_{x_1} = n_{x_2} = n_1$, $n_{y_1} = n_{y_2} = n_2$, $n_{z_1} = n_{z_2} = n_3$.

c	$(\lambda_{x_1}, \lambda_{x_2}, \lambda_{y_1}, \lambda_{y_2}, \lambda_{z_1}, \lambda_{z_2})$	(n_1, n_2, n_3)	JEL		NA	
			CP(%)	AL	CP(%)	AL
0	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	86.1	.151	85.3	.143
		(20, 25, 30)	88.9	.093	88.1	.089
		(30, 30, 30)	90.9	.080	89.2	.078
		(60, 60, 60)	91.0	.056	90.3	.055
		(80, 80, 80)	90.3	.048	89.8	.047
		(100,100,100)	89.5	.043	89.5	.042
0.25	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	86.9	.146	85.0	.138
		(20, 25, 30)	90.1	.089	88.7	.085
		(30, 30, 30)	90.9	.076	89.5	.074
		(60, 60, 60)	89.9	.053	88.5	.052
		(80, 80, 80)	90.9	.045	90.0	.045
		(100,100,100)	88.5	.040	87.4	.039
0.5	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	86.5	.135	83.8	.128
		(20, 25, 30)	89.7	.084	88.0	.081
		(30, 30, 30)	88.5	.070	86.5	.068
		(60, 60, 60)	91.4	.048	89.6	.047
		(80, 80, 80)	92.3	.041	91.7	.041
		(100,100,100)	88.2	.037	88.3	.036
0.75	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	86.8	.123	85.0	.117
		(20, 25, 30)	90.7	.073	87.4	.070
		(30, 30, 30)	89.6	.062	87.2	.060
		(60, 60, 60)	90.6	.043	89.1	.042
		(80, 80, 80)	90.8	.036	90.0	.036
		(100,100,100)	90.4	.032	89.7	.032
0.9	(5, 3, 4, 2, 4, 1)	(10, 10, 10)	87.8	.111	85.0	.107
		(20, 25, 30)	90.5	.065	87.3	.063
		(30, 30, 30)	90.9	.055	88.1	.054
		(60, 60, 60)	92.9	.038	91.2	.038
		(80, 80, 80)	91.6	.032	90.8	.032
		(100,100,100)	90.6	.029	90.0	.029

NOTE:
JEL: Jackknife Empirical Likelihood,
NA: Normal Approximation,
CP(%): Coverage Probability,
AL: Average Length.

CHAPTER 7

CONCLUSIONS

7.1 Summary

The dissertation is expected to have a broader impact on the practice of statistics and other research fields. The ROC- and AUC-type measures of diagnostic accuracy is due to the significance of diagnostic conclusions on the treatment phase of clinical trials and engineering reliability evaluations. Improvements in diagnostic accuracy will result in budgetary economy and ethical relief. In this dissertation, we focus on providing a reliable alternative in evaluating diagnostic tests with censoring through the plug-in empirical likelihood procedure.

Also, we make elaborate efforts by providing a reliable alternative in evaluating diagnostic tests through the jackknife empirical likelihood procedure. A new inference technique is constructed to compare the diagnostic treatments in discriminating three-category data. We used paired three-sample U-statistics to estimate the difference of the volumes and established the Wilk's theorem for the U-statistics rigorously. The corresponding coverage probability and average length of the confidence intervals are calculated based on the Wilk's theorem. Our JEL method for paired three-sample U-statistics is different from the existing JEL methods of univariate multi-sample U-statistics (see Jing et al. (2009) and Pan et al. (2013)). We also proposed the nonparametric normal approximation method, to make statistical inference for the difference of two volumes under the three-class ROC surfaces. The intensive simulation studies show the advantages of the JEL method over the normal approximation method in terms of coverage probability.

7.2 Future Work

In the future, we continue the study in more than one way. For example, jackknife empirical likelihood may be applied for the ROC curve with right censoring, the difference of two ROC curves with right censoring, the AUC with right censoring, and the difference of two AUC's with right censoring. In addition, we will investigate the adjusted JEL confidence intervals for the difference of two VUS's. On the other hand, we will also study the partial volume under surface (PVUS), which is another important and powerful quantity for the evaluation of the diagnostic tests. Finally, we will explore the VUS and PVUS with incomplete data.

REFERENCES

James Arvesen. Jackknifing U-statistics. *Annals of Mathematical Statistics*, 40(6):2096–2100, 1969.

Fangfang Bai and Yong Zhou. Kernel estimators of the ROC curve with censored data. *Acta Mathematicae Applicatae Sinica*, 29(1):43–54, 2013.

K. Balakrishnan. *Exponential Distribution: Theory, Methods and Applications*. CRC Press, Amsterdam, illustrated edition, 1996.

Jian Bi, Hye-Seong Lee, and Michael O'Mahony. Statistical analysis of receiver operating characteristic (ROC) curves for the ratings of then A-Not A and the same-different methods. *Journal of Sensory Studies*, 28(1):34–36, 2012.

Andrew P. Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, 30(7):1145–1159, 1997.

Gerda Claeskens, Bing-Yi Jing, Liang Peng, and Wang Zhou. Empirical likelihood confidence regions for comparison distributions and ROC curves. *Canadian Journal of Statistics*, 31 (2):173–190, 2003.

Nancy R. Cook. Use and misuse of the receiver operating characteristic curve in risk prediction. *Circulation*, 115:928–935, 2007.

Jesse Davis and Mark Goadrich. The relationship between precision-recall and ROC curves. Report, Department of Computer Science and Department of Biostatistics and Medical Informatics, University of Wisconsin Madison, 2006.

E. DeLong, D. DeLong, L. Daniel, and D. L. Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 44:837–845, 1988.

Lori E. Dodd and Margaret S. Pepe. Partial AUC estimation and regression. *Biometrics*, 59:614–623, 2003.

Richard M. Dudley. Consistency of m-estimators and one-sided bracketing. In M. Eberlein, E.and Hahn and M. Talagrand, editors, *The First International Conference on High-Dimensional Probability*, volume 43, pages 33–58. Birkhuser Basel, 1998.

L. Endrenyi, S. Fritsch, and W. Yan. c_{max}/AUC is a clearer measure than c_{max} for absorption rated in investigations of bioequivalence. *International Journal of Clinical Pharmacology, Therapy and Toxicology*, 29(10):394–399, 1991.

Tom Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27:861–874, 2006.

Thomas R Fleming and David P Harrington. *Counting processes and survival analysis*, volume 169. John Wiley and Sons, 1991.

Evarist Gine and Joel Zinn. Some limit theorems for empirical processes. *Annals of Probability*, 12(4):929–989, 1984.

David J. Hand and Robert J. Till. A simple generalization of the area under the ROC curve for multiple class classification problems. *Machine Learning*, 45:171–186, 2001.

J. A. Hanley and B. J. McNeil. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, 148(3):839–843, 1983.

N. L. Hjort, I. W. McKeague, and I. Van Keilegom. Extending the scope of empirical likelihood. *Annals of Statistics*, 37:1079–1111, 2009.

Fushing Hsieh and Bruce W. Turnbull. Nonparametric methods for evaluating diagnostic tests. *Statistica Sinica*, 6:47–62, 1996.

Xin Huang, Gengsheng Qin, Yan Yuan, and Xiao-Hua Zhou. Confidence intervals for the difference between two partial AUCs. *Australian and New Zealand Journal of Statistics*, 54(1):63–79, 2012.

B.Y. Jing, J. Yuan, and W. Zhou. Jackknife empirical likelihood. *Journal of the American Statistical Association*, 104:1224–1232, 2009.

Vladimir S. Korolyuk and IUrii V. Borovskikh. *Theory of U-Statistics*. Mathematics and Its Application. Springer, 1994.

Jeanne Kowalski and Xin M. Tu. *Modern Applied U-Statistics*. John Wiley and Sons, 2007.

Murat Kurtcephe and H. Altay Guvenir. A discretization method based on maximizing the area under receiver operating characteristic curve. *International Journal of Pattern Recognition and Artificial Intelligence*, 27(1), 2013.

D. Y. Lin and Z. Ying. A simple nonparametric estimator of the bivariate survival function under univariate censoring. *Biometrika*, 80(3):573–581, 1993.

D. Y. Lin, L. J. Wei, and Z. Ying. Checking the Cox model with cumulative sums of martingale-based residuals. *Biometrika*, 80(3):557–572, 1993.

C. X. Ling and H. Zhang. Toward Bayesian classifiers with accurate probabilities. In *the Sixth Pacific-Asia Conference on KDD*, volume 2002. Springer, 2002.

C. X. Ling, J. Huang, and H. Zhang. AUC: a statistically consistent and more discriminating measure than accuracy. In *IJCAI'03 Proceedings of the 18th international joint conference on Artificial intelligence*, pages 519–524. Morgan Kaufmann Publishers Inc, 2003.

K. Linnet. Comparison of quantitative diagnostic tests: Type I error, power, and sample size. *Statistics in Medicine*, 6(2):147–158, 1987.

C. J. Lloyd. The use of smoothed ROC curves to summarize and compare diasnostic systems. *Journal of the American Statistical Association*, 93(444):1356–1364, 1998.

C. J. Lloyd and Z. Yong. Kernel estimators for the ROC curve are better than empirical. *Statistics and Probability Letters*, 44:221–228, 1999.

Jorge M. Lobo, Alberto Jimenez-Valverde, and Real Raimundo. AUC: a misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17:145–151, 2008.

A. W. Marshall and I. Olkin. A multivariate exponential distribution. *Journal of American Statistical Association*, (62):30–41, 1967.

C. E. Metz, B. A. Herman, and Roe Cheryl A. Statistical comparison of two ROC-curve estimates obtained from partially-paired datasets. *Medical Decision Making*, 18:110–121, 1998a.

C. E. Metz, B. A. Herman, and J. H. Shen. Maximum likelihood estimation of receiver operating characteristic (ROC) curves from continuously-distributed data. *Statistics in Medicine*, 17:1033–1053, 1998b.

Douglas Mossman. Three-way ROC's. *Medical Decision Making*, 19(1):78–89, 1999.

Christos T. Nakas and Constantin T. Yiannoutsos. Ordered multiple-class ROC analysis with continuous measurements. *Statistics in Medicine*, (23):3437–3449, 2004.

Nancy A. Obuchowski. Receiver operating characteristic curves and their use in radiology. *Statistical Concepts Series*, 299:3–8, 2003.

A. B. Owen. Empirical likelihood ratio confidence intervals for a single functional. *Biometrika*, 75(2):237–249, 1988.

A. B. Owen. Empirical likelihood confidence regions. *Annals of Statistics*, 18:90–120, 1990.

A. B. Owen. Empirical likelihood for linear models. *Annals of Statistics*, 19:1725–1747, 1991.

A. B. Owen. Empirical likelihood confidence intervals for a single functional. *Biometrika*, 75:237–249, 1998.

A. B. Owen. *Empirical Likelihood.* London: Chapman and Hall, 2001.

Guangming Pan, Xiping Wang, and Wang Zhou. Nonparametric statistical inference for P(X < Y < Z). *The Indian Journal of Statistics*, 75-A:118–138, 2013.

Michael J. Pencina, Ralph B. D'Agostino Sr, and Ralph B. D'Agostino Jr. Evaluating the added predictive ability of a new marker: From area under ROC curve to reclassification and beyond. *Statistics in Medicine*, 27:157–172, 2008.

Margaret S. Pepe. A regression modeling framework for ROC curves in medical diagnostic testing. *Biometrika*, 84(3):595–608, 1997. doi: 10.1093/biomet/84.3.595.

Margaret S. Pepe. *The Statistical Evaluation of Medical Tests for Classication and Prediction*. Oxford University Press, Oxford, 2003.

Margaret S. Pepe and Tianxi Cai. The analysis of placement values for evaluating discriminatory measures. *Biometrics*, 60(2):528–535, 2004. doi: 10.1111/j.0006-341X.2004.00200.x.

G. Qin and X. Zhou. Empirical likelihood inference for the area under the ROC curve. *Biometrics*, 62:613–622, 2006.

Yongsong Qin and Lincheng Zhao. Empirical likelihood ratio confidence intervals for various differences of two populations. *System Science and Mathematics Sciences*, 13(1):23–30, 2000.

Pranab Kumar Sen. On some convergence properties of U-statistics. *Calcutta Statistical Association Bulletin*, 10:1–18, 1960.

G. R. Shorack and J. A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley and Sons Inc., New York, 1986.

M Storandt and RD Hill. Very mild senile dementia of the alzheimer type. II. psychometric test performance. *Archives of Neurology*, 46(4):383 – 386, 1989.

W. Stute and J.-L. Wang. The strong law under random censorship. *Annals of Statistics*, 21:1591–1607, 1993.

Winfried Stute. The central limit theorem under random censorship. *Annals of Statistics*, 23(2):422–439, 1995.

J. A. Swets and R.M. Pickett. *Evaluation of Diagnostic Systems: Methods from Signal Detection Theory*. Elsevier Science and Technology Books, 1982.

Lili Tian, Chengjie Xiong, Chin-Ying Lai, and Albert Vexler. Exact confidence interval estimation for the difference in diagnostic accuracy with three ordinal diagnostic groups. *Journal of Statistical Planning and Inference*, 141(1):549–558, 2011.

A.N. Tosteson and C.B. Begg. A general regression methodology for ROC curve estimation. *Medical Decision Making*, 8(3):204–215, 1988.

Janis Valeinis. *Confidence Bands for Structural Relationship Models*. Thesis, University of Goettingen, 2007.

E.S. Venkatraman and C.B. Begg. A distribution-free procedure for comparing receiver operating characteristic curves from a paired experiment. *Biometrika*, 83:835–848, 1996.

Shuwen Wan. An empirical likelihood confidence interval for the volume under ROC surface. *Statistics and Probability Letters*, 82(7):1463–1467, 2012.

J. G. Wang. A note on the uniform consistency of the Kaplan-Meier estimator. *Annals of Statististics*, 15:1313–1316, 1987.

Q.H. Wang, L.L. Yao, and P. Lai. Estimation of the area under ROC curve with censored data. *Journal of Statistical Planning and Inference*, 139:1033–1044, 2009.

Xiping Wang. *Empirical Likelihood with Applications*. Thesis, National University of Singapore, 2010.

David Wechsler. *Manual for the Wechsler Adult Intelligence Scale*, volume vi. Psychological Corporation, Oxford, England, 1955.

S. Wieand, M. Gail, and B. James. A family of nonparametric statistics for comparing diagnostic markers with paired or unpaired data. *Biometrika*, 76(3):585–592, 1989.

Chegnjie Xiong, Gerald van Belle, J. Philip Miller, and John C. Morris. Measuring and estimating diagnostic accuracy when there are three ordinal diagnostic groups. *Statistics in Medicine*, 2006(25):1251–1273, 2006.

Hanfang Yang and Yichuan Zhao. Smoothed empirical likekihood for ROC curves with censored data. *Journal of Multivariate Analysis*, 109:254–263, 2012.

Hanfang Yang and Yichuan Zhao. Smoothed jackknife empirical likelihood inference for the difference of ROC curves. *Journal of Multivariate Analysis*, 115:270–284, 2013.

Hanfang Yang and Yichuan Zhao. Smoothed jackknife empirical likelihood inference for ROC curves with missing data. *Journal of Multivariate Analysis*, 140:123–138, 2015.

Hanfang Yang, Kun Lu, and Yichuan Zhao. A nonparametric approach for partial areas under ROC curves and ordinal dominance curves. *Statistica Sinica*, 27(1), 2017.

Suqin Yao. *Empirical Likelihood Based Confidence Intervals for the Difference between Two Sensitivities of Continuous-scale Diagnostic Tests at a Fixed Level of Specificity*. Thesis, Georgia State University, 2007.

M. Zhou. Asymptotic normality of the synthetic estimator for censored survival data. *Annals of Statistics*, 20:1002–1021, 1992.

X.H. Zhou, D.K. McClish, and N.A. Obuchowski. *Statistical Methods in Diagnostic Medicine, 2nd Edition*. John Wiley and Sons, 2009.

K.H. Zou, W.J. Hall, and D.E. Shapiro. Smooth non-parametric receiver operating characteristic (ROC) curves for continuous diagnostic tests. *Statistics in Medicine*, 16:2143–2156, 1997.

Mark H. Zweig and Gregory Campbell. Receiver operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, 39(4):561–577, 1993.

Appendix A

PROOFS OF CHAPTER 2

Denote $\phi(t) = I(1 - F(t) \leq p) - R(p)$, $\bar{H}(t) = P(Y > t)$, $\tilde{H}_i(t) = P(Y > t, \delta = i)$, $i = 0, 1$, we have

$$\gamma_0(t) = exp[\int_0^{t-} \frac{d\tilde{H}_0(s)}{\bar{H}(s)}],\ C(t) = \int_0^{t-} \frac{dQ(s)}{(1 - H(s))(1 - Q(s))}.$$

Theorem A.1. *The empirical log-likelihood ratio for $R(p)$ is*

$$l(R_0(p)) = 2\sum_{j=1}^{m} \log(1 + \lambda\hat{w}_j) = 2\sum_{j=1}^{m}(\lambda\hat{w}_j - \frac{1}{2}(\lambda\hat{w}_j)^2) + R_n,$$

where

$$|R_n| \leq c\sum_{j=1}^{m} |\lambda\hat{w}_j|^3 = c|\lambda|^3 \sum_{j=1}^{m} |\hat{w}_j|^3 \leq c|\lambda|^3 \max_{1\leq j\leq m} |\hat{w}_j| \sum_{j=1}^{m} \hat{w}_j^2,$$

where $w_j = \dfrac{I(1 - F(Y_j) \leq p) - R(p)}{1 - Q(Y_j)}\eta_j$. *With Central Limit Theorem, we can prove that*

$$\frac{1}{m}\sum w_j^2 = O_p(1) < \infty.$$

Before the proof of Theorem A.1, we will first prove that

$$\Big|\frac{1}{m}\sum_{j=1}^{m} \hat{w}_j{}^2 - \frac{1}{m}\sum_{j=1}^{m} w_j^2\Big| \leq \frac{1}{m}\sum_{j=1}^{m}(\hat{w}_j - w_j)^2 + 2\Big|\frac{1}{m}\sum_{j=1}^{m}(\hat{w}_j - w_j)w_j\Big|.$$

The ROC curve is defined as

$$R_0(p) = 1 - G(F^{-1}(1 - p)) = EI(1 - F(Y) \leq p).$$

For the ROC curves with right censoring, one has

$$E(\eta) = EI(Y_0 \leq \nu) = P(Y_0 \leq \nu) = 1 - P(\nu \leq Y_0) = 1 - Q(Y_0),$$

and

$$R(p) = E\frac{I(1 - F(Y) \leq p)\eta}{1 - Q(Y)},$$

$$E\frac{I(1 - F(Y) \leq p)\eta}{1 - Q(Y)} - ER(p)\frac{\eta}{1 - Q(Y)} = 0,$$

$$E\frac{[I(1 - F(Y) \leq p) - R(p)]\eta}{1 - Q(Y)} = 0.$$

Then, we define that

$$w_j = \frac{[I(1 - F(Y_j) \leq p) - R(p)]\eta_j}{1 - Q(Y_j)},$$

$$\tilde{w_j} = \frac{[I(1 - F(Y_j) \leq p) - R(p)]\eta_j}{1 - \hat{Q}(Y_j)}, \ \hat{w}_j = \frac{[I(1 - \hat{F}(Y_j) \leq p) - R(p)]\eta_j}{1 - \hat{Q}(Y_j)},$$

where $\hat{F}$, $\hat{Q}$ are the Kaplan-Meier estimators of F, Q. Moreover, the empirical likelihood ratio for $R(p)$ is

$$\hat{R}(R(p)) = \sup\Big\{\prod_{j=1}^{m}(mp_j), \sum_{j=1}^{m} p_j = 1, p_j > 0, \sum_{j=1}^{m} \hat{w}_j p_j = 0\Big\}.$$

By the Lagrange multiplier method, we have the empirical log-likelihood ratio as

$$\hat{l}(R(p)) = -2\log\hat{R}(R(p)) = 2\sum_{j=1}^{m}\log(1 + \lambda\hat{w}_j),$$

where λ satisfies $\frac{1}{m}\sum_{j=1}^{m}\frac{\hat{w}_j}{1 + \lambda\hat{w}_j} = 0.$

Proof of Theorem A.1. By the Lagrange multiplier method, we make the following transformations:

$$\max\Big[\prod_{j=1}^{m}(mp_j)\Big] \Leftrightarrow \max\Big[\prod_{j=1}^{m}(p_j)\Big] \Leftrightarrow \max\Big[\log\prod_{j=1}^{m}(p_j)\Big] \Leftrightarrow \max\Big[\sum_{j=1}^{m}\log(p_j)\Big].$$

And we have the following conditions:

1. $p_j > 0,\ j = 1, 2, ..., m;$
2. $\sum_{j=1}^{m} p_j = 1 \Leftrightarrow \sum_{j=1}^{m} p_j - 1 = 0;$
3. $\sum_{j=1}^{m} \hat{w}_j p_j = 0.$

Define

$$H(p) = \sum_{j=1}^{m}\log(p_j) - \lambda_1\Big(\sum_{j=1}^{m} p_j - 1\Big) - \lambda_2\Big(\sum_{j=1}^{m}\hat{w}_j p_j\Big),$$

and let $\frac{\partial H(p)}{\partial p_j} = 0,\ j = 1, 2, ..., m$. Then

$$\frac{1}{p_j} - \lambda_1 - \lambda_2 w_j = 0\ \Rightarrow \lambda_1 = m.$$

And $1 - mp_j - \lambda_2 w_j p_j = 0$. Without loss of generality, we can replace λ_2/m with λ_2. Therefore,

$$mp_j = \frac{1}{1+\lambda_2 w_j},$$

then

$$R(p) = \prod_{j=1}^{m}(mp_j) = \prod_{j=1}^{m}\frac{1}{1+\lambda_2 w_j}.$$

On the other hand,

$$\sum_{j=1}^{m} w_j p_j = 0 \Rightarrow \sum_{j=1}^{m}(mp_j)w_j = 0 \Rightarrow \sum_{j=1}^{m}\frac{w_j}{1+\lambda_2 w_j} = 0.$$

□

Since we assume that $R(0) = 0$, $R(1) = 1$, and $0 \le p \le 1$. Without loss of generality, we only consider $R(p)$ for $0 < p < 1$. Moreover, we assume that $\tau_G = \sup_t\{t : G(t) = 0\}$, $\tau_F = \sup_t\{t : F(t) = 0\}$, and $\tau_G \leq \tau_F$. Then, $0 < R(p) < 1$ when $0 < p < 1$.

Lemma A.1. *Under some regularity conditions,*

$$\frac{1}{\sqrt{m}} \sum_{j=1}^{m} w_j \to N(0, \sigma^2(x)).$$

Proof of Lemma A.1. The proof can be referred to Corallary 1.2 in Stute (1995). □

Lemma A.2. *Under the same regularity conditions,*

$$\frac{1}{m} \sum_{j=1}^{m} \hat{w}_j^2 = O_p(1).$$

Proof of Lemma A.2. From the LLN,

$$\frac{1}{m} \sum_{j=1}^{m} \hat{w}_j^2 = E(w_j^2) + o_p(1) = O_p(1),$$

and

$$\begin{aligned}
\Big|\frac{1}{m} \sum_{j=1}^{m} \hat{w}_j^2 - \frac{1}{m} \sum_{j=1}^{m} w_j^2\Big| \leq& \Big|\frac{1}{m} \sum_{j=1}^{m} \hat{w}_j^2 - \frac{1}{m} \sum_{j=1}^{m} \tilde{w}_j^2\Big| + \Big|\frac{1}{m} \sum_{j=1}^{m} \tilde{w}_j^2 - \frac{1}{m} \sum_{j=1}^{m} w_j^2\Big| \\
=& D_1 + D_2,
\end{aligned}$$

where

$$\begin{aligned}
D_2 &= \Big|\frac{1}{m} \sum_{j=1}^{m} \tilde{w}_j^2 - \frac{1}{m} \sum_{j=1}^{m} w_j^2\Big| \\
&= \Big|\frac{1}{m} \sum_{j=1}^{m} (w_j - \tilde{w}_j)(w_j - \tilde{w}_j + 2\tilde{w}_j)\Big| \\
&\leq \frac{1}{m} \sum_{j=1}^{m} (w_j - \tilde{w}_j)^2 + 2\Big|\frac{1}{m} \sum_{j=1}^{m} (w_j - \tilde{w}_j)\tilde{w}_j\Big| \\
&= I_1 + I_2,
\end{aligned}$$

$$
\begin{aligned}
I_1 &= \frac{1}{m}\sum_{j=1}^{m}(w_j - \tilde{w}_j)^2 \\
&= \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{1}{1-Q(Y_j)} - \frac{1}{1-\hat{Q}(Y_j)}\Big)^2 [I(1-F(Y_j) \le p) - R(p)]^2 \eta_j^2 \\
&\le \sup_{t \le Y_{(m)}} \Big|\frac{\hat{Q}(t) - Q(t)}{1-\hat{Q}(t)}\Big|^2 \frac{1}{m}\sum_{j=1}^{m}\frac{[I(1-F(Y_j) \le p) - R(p)]^2 \eta_j^2}{[1-Q(Y_j)]^2}.
\end{aligned}
$$

By the fact in Zhou (1992), we have

$$
\sup_{t \le Y_{(m)}} \Big|\frac{\hat{Q}(t) - Q(t)}{1-\hat{Q}(t)}\Big|^2 = O_p(1).
$$

And,

$$
\frac{1}{m}\sum_{j=1}^{m}\frac{[I(1-F(Y_j) \le p) - R(p)]^2 \eta_j^2}{[1-Q(Y_j)]^2} \le \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{\eta_j}{1-Q(Y_j)}\Big)^2,
$$

where

$$
\frac{1}{m}\sum_{j=1}^{m}\Big(\frac{\eta_j}{1-Q(Y_j)}\Big)^2 = E\Big[\frac{\eta_j}{1-Q(Y_j)}\Big]^2 + o_p(1) = O_p(1).
$$

Moreover,

$$
E\Big(\frac{\eta_j}{1-Q(Y_j)}\Big) = E\Big(\frac{\eta_1}{1-Q(Y_1)}\Big) = 1.
$$

Thus $I_1 = O_p(1)$. On the other hand, $I_2 = \Big|\frac{1}{m}\sum_{j=1}^{m}(w_j - \tilde{w}_j)\tilde{w}_j\Big|$. Similarly, $I_2 = O_p(1)$. Therefore, $D_2 = I_1 + I_2 = O_p(1)$. Next, consider D_1:

$$
D_1 \le \frac{1}{m}\sum_{j=1}^{m}(\hat{w}_j - \tilde{w}_j)^2 + 2\Big|\frac{1}{m}\sum_{j=1}^{m}(\hat{w}_j - \tilde{w}_j)\tilde{w}_j\Big| = K_1 + 2K_2.
$$

Because of the fact that $|I(1-\hat{F}(Y_j) \le p) - I(1-F(Y_j) \le p)| \le 1$, it leads to that:

$$
\begin{aligned}
K_1 &= \frac{1}{m}\sum_{j=1}^{m}(\hat{w}_j - \tilde{w}_j)^2 \\
&= \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{\eta_j}{1-\hat{Q}(Y_j)}\Big)^2 [I(1-\hat{F}(Y_j) \le p) - I(1-F(Y_j) \le p)]^2 \\
&\le \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{\eta_j}{1-\hat{Q}(Y_j)}\Big)^2.
\end{aligned}
$$

Using the consistency of Kaplan-Meier estimators, we have

$$
\frac{1}{m}\sum_{j=1}^{m}\Big(\frac{\eta_j}{1-\hat{Q}(Y_j)}\Big) \Rightarrow \frac{1}{m}\sum_{j=1}^{m}\Big(\frac{\eta_j}{1-Q(Y_j)}\Big) \overset{CLT}{\Rightarrow} E\Big(\frac{\eta_1}{1-Q(Y_1)}\Big) = O_p(1).
$$

Therefore, $K_1 = O_p(1)$.

$$
\begin{aligned}
K_2 &= \Big|\frac{1}{m}\sum_{j=1}^{m}(\hat{w}_j - \tilde{w}_j)\tilde{w}_j\Big| \\
&= \Big|\frac{1}{m}\sum_{j=1}^{m}\frac{\eta_j \tilde{w}_j}{1-\hat{Q}(Y_j)}[I(1-\hat{F}(Y_j) \le p) - I(1-F(Y_j) \le p)]\Big| \\
&= O_p(1).
\end{aligned}
$$

Thus, $D_1 = K_1 + 2K_2 = O_p(1)$. Together with the fact that $D_2 = O_p(1)$, we have

$$
\Big|\frac{1}{m}\sum_{j=1}^{m}\hat{w}_j^2 - \frac{1}{m}\sum_{j=1}^{m}w_j^2\Big| \le D_1 + D_2 = O_p(1).
$$

$\square$

Proof of Theorem 2.1. $\hat{l}(R_0(p)) = 2\sum_{j=1}^{m}\log(1+\lambda\hat{w}_j)$, where λ satisfies $\sum_{j=1}^{m}\frac{\hat{w}_j}{1+\lambda\hat{w}_j} = 0$. Combining Lemma A.2 and the same argument in Owen (1991), we have $\lambda = O_p(n^{-1/2})$. By Lemma 11.2 in Owen (1991), and $Ew_j^2 < \infty$,

$$
\max_{1\le j\le m}|w_j| = o_p(n^{1/2}).
$$

$$\max_{1\le j\le m}|\tilde{w}_j| \le \max_{1\le j\le m}|\tilde{w}_j - w_j| + \max_{1\le j\le m}|w_j|,$$

where

$$\max_{1\le j\le m}|\tilde{w}_j - w_j| \le \sup_{t\le Y_{(m)}}\Big|\frac{\hat{Q}(t)-Q(t)}{1-\hat{Q}(t)}\Big|\max_{1\le j\le m}\frac{[I(1-F(t)\le p)-R(p)]\eta_j}{1-Q(t)} \le o_p(n^{1/2}),$$

where the 1st term, $\sup_{t\le Y_{(m)}}\Big|\frac{\hat{Q}(t)-Q(t)}{1-\hat{Q}(t)}\Big|$, can be proved by Zhou (1992), and the 2nd term is w_j. Then, $\max_{1\le j\le m}|\tilde{w}_j - w_j| = o_p(n^{1/2})$.

$$\max_{1\le j\le m}|\hat{w}_j| \le \max_{1\le j\le m}|\hat{w}_j - \tilde{w}_j| + \max_{1\le j\le m}|\tilde{w}_j|,$$

where

$$\begin{aligned}\max|\hat{w}_j - \tilde{w}_j| &= \max\Big|\frac{I(1-\hat{F}(Y_j)\le p) - I(1-F(Y_j)\le p)}{1-\hat{Q}(Y_j)}\eta_j\Big| \\ &= \max\Big|\frac{I(1-\hat{F}(Y_j)\le p) - I(1-F(Y_j)\le p)}{I(1-F(Y_j)\le p)-R(p)}\Big|\Big|\frac{I(1-F(Y_j)\le p)-R(p)}{1-\hat{Q}(Y_j)}\eta_j\Big|.\end{aligned}$$

The 2nd term is $\tilde{w}_j$, and the 1st term is $O_P(1)$ for a fixed $p\in(0,1)$, and $R(p)\in(0,1)$. Thus

$$\max|\hat{w}_j - \tilde{w}_j| \le O_p(1)o_p(n^{1/2}) = o_p(n^{1/2}).$$

Together with $\max|\tilde{w}_j| = o_p(n^{1/2})$, we have $\max|\hat{w}_j| = o_p(n^{1/2})$. By Lemma 11.3 in Owen (1991), and $Ew_j^2<\infty$, $\frac{1}{m}\sum_{j=1}^m|w_j|^3 = o_p(n^{1/2})$.

$$\frac{1}{m}\sum_{j=1}^m|\tilde{w}_j|^3 = \frac{1}{m}\sum_{j=1}^m(|\tilde{w}_j|^3 - |w_j|^3) + \frac{1}{m}\sum_{j=1}^m|w_j|^3.$$

$$\begin{aligned}\frac{1}{m}\sum_{j=1}^{m}(|\tilde{w}_j|^3-|w_j|^3) &= \frac{1}{m}\sum_{j=1}^{m}[(|\tilde{w}_j|-|w_j|)(\tilde{w}_j^2+w_j^2+|\tilde{w}_j||w_j|)] \\ &\leq \max|\tilde{w}_j-w_j|\Big(\frac{1}{m}\sum_{j=1}^{m}\tilde{w}_j^2+\frac{1}{m}\sum_{j=1}^{m}w_j^2+\frac{1}{m}\sum_{j=1}^{m}|\tilde{w}_j||w_j|\Big) \\ &\leq \max(\tilde{w}_j-w_j)\Big\{\frac{1}{m}\sum_{j=1}^{m}\tilde{w}_j^2+\frac{1}{m}\sum_{j=1}^{m}w_j^2+\frac{1}{m}\sqrt{\sum_{j=1}^{m}\tilde{w}_j^2\sum_{j=1}^{m}w_j^2}\Big\} \\ &= o_p(n^{1/2})(O_p(1)+O_p(1)+O_p(1)) \\ &= o_p(n^{1/2}).\end{aligned}$$

Thus $1/m\sum_{j=1}^{m}|\tilde{w}_j|^3=o_p(n^{1/2})$. Similarly, we have $\frac{1}{m}\sum_{j=1}^{m}|\hat{w}_j|^3=o_p(n^{1/2})$. From the Taylor's expansion of the empirical log-likelihood ratio,

$$\hat{l}(R(p))=2\sum_{j=1}^{m}\log(1+\lambda\hat{w}_j)=2\sum_{j=1}^{m}\Big(\lambda\hat{w}_j-\frac{1}{2}(\lambda\hat{w}_j)^2\Big)+R_n,$$

where

$$|R_n|\leq c\sum_{j=1}^{m}|\lambda\hat{w}_j|^3\leq c|\lambda|^3\sum_{j=1}^{m}|\hat{w}_j|^3=O_p(1)o_p(n^{-3/2})o_p(n^{3/2})=o_p(1).$$

On the other hand, we have

1.

$$\sum_{j=1}^{m}\frac{\hat{w}_j}{1+\lambda\hat{w}_j}=0,$$

$$\sum_{j=1}^{m}\hat{w}_j-\lambda\sum_{j=1}^{m}\hat{w}_j^2+\sum_{j=1}^{m}\frac{\lambda^2\hat{w}_j^3}{1+\lambda\hat{w}_j}=0,$$

$$\lambda=\Big(\sum_{j=1}^{m}\hat{w}_j^2\Big)^{-1}\Big(\sum_{j=1}^{m}\hat{w}_j\Big)+\Big(\sum_{j=1}^{m}\hat{w}_j^2\Big)^{-1}\Big(\sum_{j=1}^{m}\frac{\lambda^2\hat{w}_j^3}{1+\lambda\hat{w}_j}\Big),$$

where $|\hat{w}_j|\leq\max_{1\leq j\leq m}|\hat{w}_j|=o_p(n^{1/2})$. Thus $\Big(\sum_{j=1}^{m}\hat{w}_j^2\Big)^{-1}=o_p(n^{-1})$, and

$$\lambda\hat{w}_j\leq|\lambda|\sum_{j=1}^{m}|\hat{w}_j|=O_p(n^{-1/2})o_p(n^{1/2})=O_p(1).$$

We know that $1 + \lambda \hat{w}_j = O_p(1)$.

$$\lambda^2 \sum_{j=1}^{m} \frac{\hat{w}_j^3}{1+\lambda \hat{w}_j} \le O_p(n^{-1}) o_p(n^{3/2}) O_p(1) = o_p(n^{1/2}),$$

$$\Big(\sum_{j=1}^{m} \hat{w}_j^2\Big)^{-1} \Big(\sum_{j=1}^{m} \frac{\lambda^2 \hat{w}_j^3}{1+\lambda \hat{w}_j}\Big) = o_p(n^{-1/2}),$$

$$\lambda = \Big(\sum_{j=1}^{m} \hat{w}_j^2\Big)^{-1} \Big(\sum_{j=1}^{m} \hat{w}_j\Big) + o_p(n^{-1/2}).$$

2.

$$\sum_{j=1}^{m} \frac{\hat{w}_j}{1+\lambda \hat{w}_j} = 0,$$

$$\sum_{j=1}^{m} \lambda \hat{w}_j - \sum_{j=1}^{m} (\lambda \hat{w}_j)^2 + \sum_{j=1}^{m} \frac{(\lambda \hat{w}_j)^3}{1+\lambda \hat{w}_j} = 0,$$

$$\sum_{j=1}^{m} \lambda \hat{w}_j = \sum_{j=1}^{m} (\lambda \hat{w}_j)^2 - \sum_{j=1}^{m} \frac{(\lambda \hat{w}_j)^3}{1+\lambda \hat{w}_j} = \sum_{j=1}^{m} (\lambda \hat{w}_j)^2 + o_p(1),$$

because $\left|\lambda^3 \sum_{j=1}^{m} \frac{\hat{w}_j^3}{1+\lambda \hat{w}_j}\right| \le O_p(n^{-3/2}) o_p(n^{3/2}) O_p(1) = o_p(1)$. Applying Taylor's expansion, we have

$$\begin{aligned}
\frac{\sigma_1^2(p)}{\sigma^2(p)} \hat{l}(R_0(p)) &= \frac{\sigma_1^2(p)}{\sigma^2(p)} 2 \sum_{j=1}^{m} \Big(\lambda \hat{w}_j - \frac{1}{2}(\lambda \hat{w}_j)^2\Big) + O_p(1) \\
&= \frac{\sigma_1^2(p)}{\sigma^2(p)} 2 \Big[\sum_{j=1}^{m} (\lambda \hat{w}_j)^2 + o_p(1)\Big] - \frac{\sigma_1^2(p)}{\sigma^2(p)} \sum_{j=1}^{m} (\lambda \hat{w}_j)^2 + o_p(1) \\
&= \frac{\sigma_1^2(p)}{m^{-1} \sum_{j=1}^{m} \hat{w}_j^2} \frac{1}{m \sigma^2(p)} \left(\sum_{j=1}^{m} \hat{w}_j\right)^2 + o_p(1) \\
&= \frac{\sigma_1^2(p)}{m^{-1} \sum_{j=1}^{m} \hat{w}_j^2} \left[\frac{\sum_{j=1}^{m} \hat{w}_j}{\sqrt{m}\sigma(p)}\right]^2 + o_p(1)
\end{aligned}$$

and

$$\lim \frac{\sigma_1^2(p)}{m^{-1}\sum_{j=1}^{m}\hat{w}_j^2} = 1,$$

$$\frac{1}{\sqrt{m}\sigma(p)}\sum_{j=1}^{m}\hat{w}_j \xrightarrow{\mathcal{D}} N(0,1),$$

$$\left[\frac{1}{\sqrt{m}\sigma(p)}\sum_{j=1}^{m}\hat{w}_j\right]^2 \to \chi_1^2.$$

Thus, $\hat{l}(R_0(p)) \xrightarrow{D} \gamma(R_0(p))\chi_1^2$, as $m \to \infty$. $\square$

Appendix B

PROOFS OF CHAPTER 3

We observe $(X_{1i}, X_{2i}, \xi_{1i}, \xi_{2i})$, $i = 1, 2, ..., n$, where

$$X_{1i} = \min(X_{1i}^0, U_i),\ \xi_{1i} = I(X_{1i}^0 \leq U_i);\ X_{2i} = \min(X_{2i}^0, U_i),\ \xi_{2i} = I(X_{2i}^0 \leq U_i);$$

and similarly, we observe $(Y_{1j}, Y_{2j}, \eta_{1j}, \eta_{2j})$, $j = 1, 2, ..., m$, where

$$Y_{1j} = \min(Y_{1j}^0, V_j),\ \eta_{1j} = I(Y_{1j}^0 \leq V_j);\ Y_{2j} = \min(Y_{2j}^0, V_j),\ \eta_{2j} = I(Y_{2j}^0 \leq V_j);$$

where $I(\cdot)$ denotes the indicator function. Denote

$$\hat{w}_{1i} = \frac{[I(1 - \hat{F}_1(X_{2i}) \leq p) - R_1(p)]\xi_{2i}}{1 - \hat{Q}_1(X_{2i})},\ \hat{w}_{2j} = \frac{[I(1 - \hat{F}_2(Y_{2j}) \leq p) - R_1(p) + D]\eta_{2j}}{1 - \hat{Q}_2(Y_{2j})},$$

$$\Delta = D(p) = R_1(p) - R_2(p).$$

Let $\theta = R_1(p) = ROC_1(p)$, and

$$\hat{\alpha}_{1i} = \frac{\partial \hat{w}_{1i}}{\partial \theta} = -\frac{\xi_{2i}}{1 - \hat{Q}_1(X_{2i})},\ \hat{\alpha}_{2j} = \frac{\partial \hat{w}_{2j}}{\partial \theta} = -\frac{\eta_{2j}}{1 - \hat{Q}_2(Y_{2j})}.$$

Refer to Assumption 4 on page 36 of Valeinis (2007). (a) For the 1st sample $X_1, ...X_n$, denote $R_1(p) = \theta$, we make the following assumptions:

(A1) $E_F w_1^2(X, \theta, \hat{h}) > 0$, $\alpha_1(X, \theta, \hat{h})$ is continuous in a neighborhood of θ_0, $\alpha_1(X, \theta, \hat{h})$, $w_1^3(X, \theta, \hat{h}) > 0$ are bounded by some integrable function $G_1(X)$ in this neighborhood, $E_F \alpha_1(X, \theta, \hat{h})$ is nonzero.

(A2) For some subset $\bar{\mathcal{H}}$ of $\mathcal{H}$ such that $P\{\hat{h} \in \mathcal{H}\} \to 1$, and for some $\eta \in (1/3, 1/2)$, the class functions $\mathcal{F} = \{w_1(., \theta, h) : |\Delta| = |\theta - \theta_0| \leq cn^{-\eta}, h \in \bar{\mathcal{H}}\}$ with a positive constant

$c < \infty$ has the strong Gilvenko-Cantelli property with the almost sure convergence rate

$$\sup_{|\Delta| \le cn^{-\eta}} \Big| \frac{1}{n} \sum_{i=1}^{n} \{w_1(X_i, \theta, h) - Ew_1(X, \theta, h)\} \Big| = O(\beta_1),$$

where $\beta_1 = o(n^{-\eta})$.

(A3) For the functions $w_1^2(X, \theta, h)$, $w_1^3(X, \theta, h)$ and $\alpha_1(X, \theta, h)$ we assume that the strong Gilvenko-Cantelli property holds, i.e., the above equation holds.

(A4) Assume that $Ew_1(X, \theta_0, \hat{h})$ a.s., where $\beta_2 = o(n^{-\eta})$.

(b) Assume that for the 2nd sample $Y_1, ...Y_m$, (A1) - (A4) also hold for the functions $w_2(Y, \theta, h)$, $w_2^2(Y, \theta, h)$, $w_2^3(Y, \theta, h)$ and $\alpha_2(Y, \theta, h)$.

Proof of (A1). 1. ($E\hat{w}_1^2 > 0$.)

$$\begin{aligned} E\hat{w}_1^2 =& E\Big\{ \frac{[I(1 - \hat{F}_1(X_2) \le p) - R_1(p)]\xi_2}{1 - \hat{Q}_1(X_2)} \Big\}^2 \\ =& E\Big\{ \frac{[I(1 - \hat{F}_1(X_2) \le p) - R_1(p)]^2}{[1 - \hat{Q}_1(X_2)]^2} I(X_1^0 \le U_1) \Big\} \\ =& E\Big\{ \frac{[I(1 - \hat{F}_1(X_2) \le p) - R_1(p)]^2}{[1 - \hat{Q}_1(X_2)]^2} | X_1^0 \le U_1 \Big\}. \end{aligned}$$

$P(X_1 \le U_1) > 0$, and

$$E\Big(E\Big\{ \frac{[I(1 - \hat{F}_1(X_2) \le p) - R_1(p)]^2}{[1 - \hat{Q}_1(X_2)]^2} \Big| X_1^0 \le U_1 \Big\} \Big) = \frac{[(I(1 - \hat{F}_1(X_2)) \le p) - R_1(p)]^2}{[1 - \hat{Q}_1(X_2)]^2} > 0.$$

Therefore $E\hat{w}_1^2 > 0$.

2. ($\hat{\alpha}_1$ is continuous in a neighbourhood of θ_0.)

$\hat{\alpha}_1$ is a constant for $R_1(p)$, then $\hat{\alpha}_1$ is continuous in a neighbourhood of $\theta_0 = R_1(p)$.

3. (α_1 and w_1^3 are bounded by some integrable function $G_1(X)$ in this neighbourhood.)

4. ($E\hat{\alpha}_1$ is nonzero.)

$$E\hat{\alpha}_1 = E\Big\{ \frac{-1}{1 - \hat{Q}_1(X_2)} | X_1 \le U_1 \Big\}.$$

$P(X_1 \leq U_1) > 0$, and $\dfrac{-1}{1-\hat{Q}_1(X_2)} < 0$, then $E\hat{\alpha}_1 < 0 (\neq 0)$. $\square$

Proof of (A2). Let $\bar{H} = \{F_1, Q_1\}$. $P(\hat{F}_1 = F_1, \hat{Q}_1 = Q_1) \to 1$, by Wang (1987). We have

$$\begin{aligned}
E\frac{I(1-\hat{F}_1(X_2) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_2)}\xi_2 =& E\frac{I(1-\hat{F}_1(X_2) \leq p) - R_1(p)}{1-\hat{Q}_1(X_2)}\xi_2 - E\Big(\frac{\Delta\xi_{2i}}{1-Q_1(X_2)}\Big) \\
=& 0 - \Delta E(\frac{\xi_{2i}}{1-Q_1(X_2)}) \\
=& -\Delta.
\end{aligned}$$

For $\eta \in (1/3, 1/2)$, $c < \infty$,

$$\begin{aligned}
&\sup_{|\Delta| \leq cn^{-\eta}} \Big| \frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - E\frac{I(1-\hat{F}_1(X_2) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_2)}\xi_{2i}\Big| \\
=& \sup_{|\Delta| \leq cn^{-\eta}} \Big| \frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_{2i})}\xi_{2i} + \Delta\Big| \\
=& \sup_{|\Delta| \leq cn^{-\eta}} \Big| \frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i} + \frac{1}{n}\sum_{i=1}^{n}\Delta\Big(1 - \frac{\xi_{2i}}{1-Q_1(X_{2i})}\Big)\Big| \\
\leq& \Big| \frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big| + \sup_{|\Delta| \leq cn^{-\eta}} |\Delta| \Big| \frac{1}{n}\sum_{i=1}^{n}\Big(1 - \frac{\xi_{2i}}{1-Q_1(X_{2i})}\Big)\Big| \\
\triangleq& D_1 + D_2.
\end{aligned}$$

By Marcinkiewicz-Zygmund strong law of large number (SLLN), one has

$$\begin{aligned}
&\frac{1}{n^{p_1}}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - E\Big(\frac{1}{n^{p_1}}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big) \\
=& \frac{1}{n^{p_1}}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - 0 \\
\to& 0 \ a.s.
\end{aligned}$$

for $\frac{1}{2} < p_1 \leq 1$.

$$n^{1-p_1}\frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i} = o(1) \ a.s.$$

for $0 \leq 1-p_1 < \frac{1}{2}$. That is,

$$D_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i} = o(n^{p_1-1})\ a.s.$$

for $0 \leq 1-p_1 < \frac{1}{2}$. On the other hand, we have

$$D_2 = \sup_{|\Delta|\leq cn^{-\eta}}|\Delta|\Big|\frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_{2i}}{1-Q_1(X_{2i})})\Big| \leq cn^{-\eta}\Big|\frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_{2i}}{1-Q_1(X_{2i})})\Big|.$$

By the SLLN, $\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-Q_1(X_{2i})}-1\right| = o(1)$ a.s. and $cn^{-\eta} = O(n^{-\eta})$. Thus $D_2 \leq O(n^{-\eta})o(1) = o(n^{-\eta})$. That is, $D_2 = O(\beta_0)$, where $\beta_0 = o(n^{-\eta})$. Set $\eta_0 = \eta \in (1/3, 1/2) \subset (0, 1/2)$, then $D_1 = o(n^{\eta})$. Thus,

$$\sup_{|\Delta|\leq cn^{-\eta}}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{I(1-\hat{F}_1(X_{2i}) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - E\frac{I(1-\hat{F}_1(X_2) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_2)}\xi_2\Big| = O(\beta_1),$$

where $\beta_1 = o(n^{-\eta})$. Thus we proved that

$$w_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{I(1-\hat{F}_1(X_2) \leq p) - R_1(p) - \Delta}{1-\hat{Q}_1(X_2)}$$

has the strong Gilvenko-Cantelli property. □

Proof of (A3). (For $p \in (0,1)$, w_1 is bounded. By Dudley (1998), both w_1^2 and w_1^3 have the strong Gilvenko-Cantelli property.)

g is bounded, say, $\|g\|_\infty < \infty$, i.e., $\sup_x |g(x)| = \|g\|_\infty < \infty$. Denote

$$g_1 = g_1^* \cdot I_{[-M,M]}(\cdot),\ g_1^* = x^2,\ g_2 = g_2^* \cdot I_{[-M,M]}(\cdot),\ g_2^* = x^3.$$

Since w_1^3 is assumed to be bounded, w_1 is bounded as well. Suppose $w_1^3(X, \theta, \hat{h})$ is bounded by an integrable function $G_1(x)$ in the neighbourhood of θ_0, w_1 is bounded by $G_1^{\frac{1}{3}}(x)$. By

Stute and Wang (1993), we have

$$\sup_x |\hat{F}_1(x) - F_1(x)| \to 0 \; a.s.; \; \sup_x |\hat{Q}_1(x) - Q_1(x)| \to 0 \; a.s.$$

Assume that $h_0 \in \mathcal{H}$, $\hat{h} \in \bar{H} \subset \mathcal{H}$ and $t \in \mathcal{T}$, where $\mathcal{T}$ is some interval. Then $\forall h$, $\hat{h} \to h$, we have $w_1^3(X, \theta, \hat{h}) \leq G_1(x)$. Thus $w_1^3(X, \theta, h) \leq G_1(x)$. If $G(x)$ is integrable, then $G(x)$ is bounded $a.s.$ Thus $w_1^3(X, \theta, h)$ is bounded $a.s.$

$\alpha_1 = -\dfrac{\xi_2}{1 - \hat{Q}_1(X_2)}$ does not vary as $R_1(p)$ or Δ. Then

$$\sup_{|\Delta| \leq cn^{-\eta}} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{-\xi_{2i}}{1 - Q_1(X_{2i})} - E \frac{\xi_2}{1 - \hat{Q}_1(X_2)} \right| = -\frac{1}{n} \sum_{i=1}^{n} \frac{\xi_{2i}}{1 - Q_1(X_{2i})} + 1.$$

By Marcinkiewicz-Zygmund strong law of large number (SLLN),

$$1 - \frac{1}{n} \sum_{i=1}^{n} \frac{\xi_{2i}}{1 - Q_1(X_{2i})} = O(\beta_1'),$$

where $\beta_1' = o(n^{-\eta_2})$, $\eta_2 \in (1/3, 1/2)$. Then w_1, w_1^2, w_1^3, and α_1 have the strong Gilivenko-Cantelli property. $\square$

Proof of (A4). Now we rewrite $\hat{E}w_1$ as

$$E\hat{w}_1 = \Big(E\hat{w}_1 - \frac{1}{n} \sum_{i=1}^{n} \hat{w}_{1i}\Big) + \Big(\frac{1}{n} \sum_{i=1}^{n} \hat{w}_{1i} - \frac{1}{n} \sum_{i=1}^{n} w_{1i}\Big) + \Big(\frac{1}{n} \sum_{i=1}^{n} w_{1i}\Big) \triangleq E_1 + E_2 + E_3.$$

By Marcinkiewiz-Zygmund strong law of large number,

$$|E_1| = \left| E\hat{w}_1 - \frac{1}{n} \sum_{i=1}^{n} \hat{w}_{1i} \right| = O(\beta_2^{(1)}),$$

where $\beta_2^{(1)} = o(n^{-\eta_3^{(1)}})$, $\eta_3^{(1)} \in (1/3, 1/2)$.

$$|E_3| = \left| \frac{1}{n} \sum_{i=1}^{n} w_{1i} \right| = \left| 0 - \frac{1}{n} \sum_{i=1}^{n} w_{1i} \right| = \left| Ew_1 - \frac{1}{n} \sum_{i=1}^{n} w_{1i} \right| = O(\beta_2^{(2)}),$$

where $\beta_2^{(2)} = o(n^{-\eta_3^{(2)}})$, $\eta_3^{(2)} \in (1/3, 1/2)$. Next, we define

$$w_{1i} = \frac{I(1 - F_1(X_{2i}) \leq p) - R_1(p)}{1 - Q_1(X_{2i})}\xi_{2i};$$

$$\hat{w}_{1i} = \frac{I(1 - \hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1 - \hat{Q}_1(X_{2i})}\xi_{2i}, \;\; \tilde{w}_{1i} = \frac{I(1 - \hat{F}_1(X_{2i}) \leq p) - R_1(p)}{1 - Q_1(X_{2i})}\xi_{2i}.$$

Then, E_2 can be rewritten as

$$E_2 = \frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n w_{1i} = \frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} + \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^n w_{1i},$$

where

$$\begin{aligned}
\Big|\frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^n w_{1i}\Big| =& \Big|\frac{1}{n}\sum_{i=1}^n \frac{I(1 - \hat{F}_1(X_{2i}) \leq p) - I(1 - F_1(X_{2i}) \leq p)}{1 - Q_1(X_{2i})}\xi_{2i}\Big| \\
\leq& \frac{1}{n}\sum_{i=1}^n \frac{|I(1 - \hat{F}_1(X_{2i}) \leq p) - I(1 - F_1(X_{2i}) \leq p)|}{1 - Q_1(X_{2i})}\xi_{2i} \\
\leq& \max_i |I(1 - \hat{F}_1(X_{2i}) \leq p) - I(1 - F_1(X_{2i}) \leq p)|\frac{1}{n}\sum_{i=1}^n \frac{\xi_{2i}}{1 - Q_1(X_{2i})}.
\end{aligned}$$

By the law of large numbers,

$$\frac{1}{n}\sum_{i=1}^n \frac{\xi_{2i}}{1 - Q_1(X_{2i})} = E\frac{\xi_2}{1 - Q_1(X_2)} + o(1),$$

and we have

$$\begin{aligned}
&\max_i |I(1 - \hat{F}_1(X_{2i}) \leq p) - I(1 - F_1(X_{2i}) \leq p)| \\
=& \max_i |I(\hat{F}_1(X_{2i}) \geq 1 - p) - I(F_1(X_{2i}) \geq 1 - p)| \\
=& \max_i I\{\min_i[\hat{F}_1(X_{2i}, F_1(X_{2i}] < 1 - p \leq \max_i[\hat{F}_1(X_{2i}), F_1(X_{2i})]\}.
\end{aligned}$$

By Stute and Wang (1993), $\sup_x |\hat{F}_1(x) - F_1(x)| \to 0$, *a.s.*, i.e.,

$$\sup_x [\max(\hat{F}_1(x), F_1(x)) - \min(\hat{F}_1(x), F_1(x))] \to 0, \ a.s.$$

Then we have

$$\max_i I\{\min_i [\hat{F}_1(X_{2i}, F_1(X_{2i}] < 1 - p \le \max_i [\hat{F}_1(X_{2i}), F_1(X_{2i})]\} = 0, \ a.s.$$

Therefore,

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n} w_{1i} = 0, \ a.s.$$

Next, consider $\frac{1}{n}\sum_{i=1}^{n} \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{1i}$. We know that $\frac{1}{n}\sum_{i=1}^{n} w_{1i} = O(\beta_2^{(2)})$,

$$\frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{1i} = O(\beta_2^{(5)}), \ a.s.,$$

where $\beta_2^{(5)} = o(n^{-\eta_3^{(5)}})$, $\eta_3^{(2)} \in (1/3, 1/2)$. Thus,

$$\begin{aligned}
&\Big|\frac{1}{n}\sum_{i=1}^{n} \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{1i}\Big| \\
=&\frac{1}{n}\sum_{i=1}^{n}\Big\{\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - \frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-Q_1(X_{2i})}\xi_{2i}\Big\} \\
\le& \sup_{s \le X_(n)} \Big|\frac{\hat{Q}_1(s) - Q_1(s)}{1-\hat{Q}_1(s)}\Big| \Big|\frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-Q_1(X_{2i})}\xi_{2i}\Big|,
\end{aligned}$$

where

$$\sup_{s \le X_(n)} \Big|\frac{\hat{Q}_1(s) - Q_1(s)}{1-\hat{Q}_1(s)}\Big| = O(1),$$

$$\Big|\frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{1i}\Big| = \Big|\frac{1}{n}\sum_{i=1}^{n} \frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-Q_1(X_{2i})}\xi_{2i}\Big| = O(\beta_2^{(5)}).$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n} \tilde{w}_{1i} = O(\beta_2^{(5)}).$$

Therefore,

$$\begin{aligned}E_2 =&\frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i}-\frac{1}{n}\sum_{i=1}^{n}w_{1i}\\=&\frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i}-\frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{1i}+\frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{1i}-\frac{1}{n}\sum_{i=1}^{n}w_{1i}\\=&O(\beta_2^{(6)}),\end{aligned}$$

where $\beta_2^{(6)} = o(n^{-\eta_3^{(6)}})$, $\eta_3^{(6)} \in (1/3, 1/2)$. Moreover,

$$E\hat{w}_1 = E_1 + E_2 + E_3 = O(\beta_2),$$

where $\beta_2 = o(n^{-\eta_3})$, $\eta_3 \in (1/3, 1/2)$. □

Thus Assumption 4 in Valeinis (2007) is proved. Next consider Assumption 5. (a) For the 1st sample $X_1, ...X_n$, we make the following assumptions:

(B1) $n^{1/2}M_{1n}(\theta_0, t, \hat{h}) \xrightarrow{d} U_1(t)$, where $U_1(t) \sim N(0, V_1(t))$.

(B2) $\sup_{t\in\mathcal{T}} |S_{1n}(\theta_0, t, \hat{h}) - V_2(t)| \xrightarrow{p} 0$.

(B3) $\sup_{t\in\mathcal{T}} |n^{-1}\sum_{i=1}^{n}\alpha(X_i, \theta_0, \hat{h}) - V_3(t)| \xrightarrow{p} 0$.

(B4) $w_1^3(X, \theta_0, \hat{h})$ is bounded by some integrable function $G_{11}(X)$.

(B5) For $R_1^*(p) \in (R_1(p) - c_0n^{-\eta}, R_1(p) + c_0n^{-\eta})$, where $c_0 > 0$, $\theta_0 = R_1(p)$, $\theta = R_1^*(p)$

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial S_{1n}(R_1^*(p), t, \hat{h})}{\partial R_1^*(p)} = O_p(1);$$

$$\frac{1}{n}\sum_{i=1}^{n}\alpha(X_i, \theta, \hat{h}) = O_p(1);$$

$$\frac{1}{n}\sum_{i=1}^{n}w_1^3(X_i, \theta, \hat{h}) = O_p(1).$$

(b) Assume that for the 2nd sample $Y_1, ...Y_m$, (B1)-(B5) also hold for the functions $w_2(Y, \theta, h)$, $w_2^2(Y, \theta, h)$, $w_2^3(Y, \theta, h)$ and $\alpha_2(Y, \theta, h)$, and $S_{2m}(\theta, t, h)$ with functions $M_1(t)$, $M_2(t)$ and $M_3(t)$ instead of $V_1(t)$, $V_2(t)$ and $V_3(t)$.

(B1) and (B2) can be proved similar to the one sample case, that is, empirical confidence intervals for the ROC curves with right censored data.

Proof of (B3). $\hat{Q}_1(X_{2i})$ is the Kaplan-Meier (K-M) estimator of Q, then $\hat{Q}_1(X_2) \xrightarrow{\mathcal{P}} Q_1(X_2)$ point-wise, and

$$\frac{\xi_2}{1-\hat{Q}_1(X_2)} \xrightarrow{\mathcal{D}} \frac{\xi_2}{1-Q_1(X_2)}.$$

By the law of large number,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_2}{1-\hat{Q}_1(X_2)} = -E\frac{\xi_2}{1-Q_1(X_2)} + o_p(1).$$

Assume that $\hat{\alpha} = \dfrac{\xi_2}{1-\hat{Q}_1(X_2)}$ is bounded by an absolute integrable function, then

$$E\frac{\xi_2}{1-\hat{Q}_1(X_2)} \to E\frac{\xi_2}{1-Q_1(X_2)} = 1,$$

by the Lebesgue's Dominated Convergence Theorem. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_{2i}}{1-\hat{Q}_1(X_{2i})} = -1 + o_p(1),$$

that is,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_{2i}}{1-\hat{Q}_1(X_{2i})} \xrightarrow{\mathcal{P}} -1.$$

□

Proof of (B4). The bounded conditions in (B4) can be directly proved by the assumptions as those in (A1). □

Proof of (B5). (1) For $R_1^*(p) \in (R_1(p) - c_0 n^{-\eta}, R_1(p) + c_0 n^{-\eta})$, $c_0 > 0$,

$$
\begin{aligned}
&\Big|\frac{\partial}{\partial R_1^*(p)} \frac{1}{n} \sum_{i=1}^{n} \Big[\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)}{1-\hat{Q}_1(X_{2i})}\Big]^2 \xi_{2i}\Big| \\
=&\Big|\frac{1}{n} \sum_{i=1}^{n} \Big\{2 \frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)}{1-\hat{Q}_1(X_{2i})} \frac{-1}{1-\hat{Q}_1(X_{2i})} \xi_{2i}\Big\}\Big| \\
=&2\Big|\frac{1}{n} \sum_{i=1}^{n} \Big\{\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)}{1-\hat{Q}_1(X_{2i})} \xi_{2i} \frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big\}\Big| \\
\le&2 \max_i \Big|\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)}{1-\hat{Q}_1(X_{2i})} \xi_{2i}\Big| \Big|\frac{1}{n} \sum_{i=1}^{n} \frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big| \\
\triangleq& S_1 + S_2,
\end{aligned}
$$

where

$$
S_1 = 2 \max_i \Big|\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)}{1-\hat{Q}_1(X_{2i})} \xi_{2i}\Big|,
$$

$$
S_2 = \Big|\frac{1}{n} \sum_{i=1}^{n} \frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big|.
$$

By the conclusion in (B3),

$$
S_2 = \Big|\frac{1}{n} \sum_{i=1}^{n} \frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big| \xrightarrow{\mathcal{P}} 1,
$$

thus $S_2 = O_p(1)$.

$$
S_1 \le 2 \max_i |I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)| \max_i \Big|\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big|,
$$

where $\max_i \Big|\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big|$ is bounded by a constant $c_{00} \in R$ as given in the assumptions in

(B4) and (A1).

$$
\begin{aligned}
&2\max_i |I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)| \\
=&2\max_i |I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p) - cn^{-\eta}| \\
\le&2\{\max_i |I(1-\hat{F}_1(X_{2i}) \le p)| + \max |R_1(p)| + \max |cn^{-\eta}|\} \\
\le&2(1+1+c_0 n^{-\eta}) \\
=&4+2c_0 n^{-\eta},
\end{aligned}
$$

where $c \in (-c_0, c_0)$, $\eta \in (1/3, 1/2)$. Thus, $S_1 \le c_{00} = 4+2c_0 n^{-\eta}$, i.e., $S_1 = O_p(1)$. Therefore,

$$
\Big|\frac{\partial}{\partial R_1^*(p)} \frac{1}{n}\sum_{i=1}^{n}\Big[\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)}{1-\hat{Q}_1(X_{2i})}\Big]^2 \xi_{2i}\Big| = O_p(1).
$$

(2) By the conclusions in (B3), we have $\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big| = O_p(1)$.

(3)

$$
\begin{aligned}
\Big|\frac{1}{n}\sum_{i=1}^{n}\hat{w}_1^3\Big| =&\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big)^3\Big| \\
\le&\frac{1}{n}\sum_{i=1}^{n}\Big|\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big|^3 \\
\le&\max_i |I(1-\hat{F}_1(X_{2i}) \le p) - R_1^*(p)|^3 \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big)^3 \\
\le&(\max_i I(1-\hat{F}_1(X_{2i}) \le p) + \max_i R_1(p) + c_0 n^{\eta})^3 \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big)^3 \\
\le&(2+c_0 n^{-\eta})^3 \frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big)^3 \\
\le&(2+c_0 n^{-\eta})^3 \Big[\max_i \frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big]^2 \frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})},
\end{aligned}
$$

where $c \in (-c_0, c_0)$, $\eta \in (1/3, 1/2)$. We know that $\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}$ is bounded and

$\frac{1}{n}\sum_{i=1}^n \frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})} \xrightarrow{\mathcal{P}} 1$ from (B3), then

$$\frac{1}{n}\hat{w}_1^3 = \frac{1}{n}\sum_{i=1}^n \Big(\frac{I(1-\hat{F}_1(X_{2i}) \le p) - R_1(p)}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big)^3 = O_p(1).$$

$\square$

As in the proof of (A3), we know that $|w_1^3(x,\theta_0,\hat{h})| \le M_{00}$, $|\alpha_1(x,\theta_0,\hat{h})| \le M_{00}$ in probability. Similarly, $|w_2^3(x,\theta_0,\hat{h})| \le M_{00}$, $|\alpha_2(x,\theta_0,\hat{h})| \le M_{00}$ in probability. For $\lambda_1' \in [0,\lambda_1^0]$, $\lambda_2' \in [0,\lambda_2^0]$, and $\theta \in [\theta_0 - cn^\eta, \theta_0 + cn^\eta]$,

$$\frac{\partial^2 Q_{1,n}(\theta,0,0)}{\partial\theta^2} = \frac{\partial^2 Q_{2,n}(\theta,0,0)}{\partial\theta^2} = 0.$$

$$\frac{\partial^2}{\partial\lambda_1^2}Q_{1,n}(\theta,0,0)\Big|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}} = 2\frac{1}{n}\sum_{i=1}^n \Big[\frac{w_1(x,\theta_0,\hat{h})}{1+\lambda_1' w_1(x,\theta_0,\hat{h})}\Big]^3.$$

When $\lambda_1' = 0$,

$$\frac{\partial^2}{\partial\lambda_1^2}Q_{1,n}(\theta,0,0)\Bigg|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}} = 2\frac{1}{n}\sum_{i=1}^n w_1(x,\theta_0,\hat{h})^3 < \infty,$$

in probability. When $\lambda_1' = \lambda_1^0$, $\frac{1}{n}\sum_{i=1}^n \frac{w_1(x,\theta_0,\hat{h})}{1+\lambda_1' w_1(x,\theta_0,\hat{h})} = 0$, and

$$\frac{1}{n}\sum_{i=1}^n \Big[\frac{w_1(x,\theta_0,\hat{h})}{1+\lambda_1' w_1(x,\theta_0,\hat{h})}\Big]^3 < \infty,$$

in probability. Given $\theta = \theta_0$, and $\lambda_2 = 0$, we know that $\frac{\partial^2}{\partial\lambda_1^2}Q_{1,n}(\theta,0,0)\Big|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}}$ is a continuous function of λ_1'. Thus for $\lambda_1' \in [0,\lambda_1^0]$, $\Big|\frac{\partial^2}{\partial\lambda_1^2}Q_{1,n}(\theta,0,0)|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}}\Big| < M$ i.e.,

$\frac{\partial^2}{\partial\lambda_1^2}Q_{1,n}(\theta,0,0)\Big|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}} = O_p(1)$. Similarly, $\frac{\partial^2}{\partial\lambda_2^2}Q_{2,n}(\theta,0,0)\Big|_{\substack{\theta=\theta_0,\\ \lambda_1=0,\\ \lambda_2=\lambda_2'}} = O_p(1)$ for $\lambda_2' \in [0,\lambda_2^0]$.

$$\frac{\partial^2}{\partial\lambda_1^2}Q_{3,n}(\theta,0,0)\Bigg|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}} = -\frac{2}{n}\sum_{i=1}^{n}\alpha_1(x,\theta_0,\hat{h})\frac{w_1(x,\theta_0,\hat{h})}{[1+\lambda_1'w_1(x,\theta_0,\hat{h})]^3}.$$

When $\lambda_1' = 0$,

$$\left|\frac{\partial^2}{\partial\lambda_1^2}Q_{3,n}(\theta,0,0)\Big|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}}\right| = \left|-\frac{2}{n}\sum_{i=1}^{n}\alpha_1(x,\theta_0,\hat{h})w_1(x,\theta_0,\hat{h})\right| < M.$$

When $\lambda_1' = \lambda_1^0$,

$$\left|\frac{\partial^2}{\partial\lambda_1^2}Q_{3,n}(\theta,0,0)\Big|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}}\right| = \left|-\frac{2}{n}\sum_{i=1}^{n}\alpha_1(x,\theta_0,\hat{h})\frac{w_1(x,\theta_0,\hat{h})}{[1+\lambda_1'w_1(x,\theta_0,\hat{h})]^2}\right| < M.$$

$\frac{\partial^2}{\partial\lambda_1^2}Q_{3,n}(\theta,0,0)$ is continuous for λ_1', thus for $\lambda_1' \in [0,\lambda_1^0]$,

$$\frac{\partial^2}{\partial\lambda_1^2}Q_{3,n}(\theta,0,0)\Bigg|_{\substack{\theta=\theta_0,\\ \lambda_1=\lambda_1',\\ \lambda_2=0}} = O_p(1).$$

Similarly,

$$\frac{\partial^2}{\partial\lambda_2^2}Q_{3,n}(\theta,0,0)\Bigg|_{\substack{\theta=\theta_0,\\ \lambda_1=0,\\ \lambda_2=\lambda_2'}} = O_p(1).$$

Therefore, Assumption 6 holds, and therefore Assumptions 3-6 hold. By Theorem 21 in Valeinis (2007),

$$-2\log R(\Delta,\theta,t,\hat{h}) \xrightarrow{\mathcal{D}} \frac{V_3^2(t)M_2(t)+kM_3^2(t)V_2(t)}{M_1(t)V_3(t)^2+kV_1(t)M_3^2(t)}\chi_1^2.$$

Appendix C

PROOFS OF CHAPTER 4

Theorem C.1 (Theorem 2.1 in Hjort et al. (2009))**.** *Suppose that*

(A0) $P\{EL_n(\theta_0, \hat{h}) \to 0\}$.

(A1) $\Sigma_{i=1}^n m_n(X_i, \theta_0, \hat{h}) \xrightarrow{\mathcal{D}} U$.

(A2) $a_n \Sigma_{i=1}^n m_n^{\otimes 2}(X_i, \theta_0, \hat{h}) \xrightarrow{pr} V_2$.

(A3) $a_n \max_{1\le i\le n} \|m_n(X_i, \theta_0, \hat{h})\| \xrightarrow{pr} 0$.

If (A0)-(A3) hold, then

$$-2a_n^{-1} \log EL_n(\theta_0, \hat{h}) \xrightarrow{\mathcal{D}} U^T V_2^{-1} U.$$

We will show (A0)-(A3) hold first. Thus Theorem 4.2 can be proved using Theorem C.1. Here we prove that (A1) holds in the first place.

Lemma C.1.

$$\frac{1}{\sqrt{m}} \sum_{j=1}^m \hat{Z}_j = \frac{1}{\sqrt{m}} \sum_{j=1}^m \frac{(\hat{F}(Y_j) - \Delta)\eta_j}{1 - \hat{Q}(Y_j)} \to N(0, \sigma^2).$$

Proof of C.1. Let

$$Z_j = \frac{(F(Y_j) - \Delta)\eta_j}{1 - Q(Y_j)},$$

then replace F and Q with their Kaplan-Meier estimators,

$$\hat{Z}_j = \frac{(\hat{F}(Y_j) - \Delta)\eta_j}{1 - \hat{Q}(Y_j)},$$

that is, let $m_n(X_i, \theta_0, \hat{h})$ be the estimate of $m(X_i, \theta_0, h)$. $m_{n_j} = \dfrac{1}{\sqrt{m}}\hat{Z}_j$, and

$$
\begin{aligned}
\sum_{j=1}^{m} m_{n_j} - \Delta &= \frac{1}{\sqrt{m}} \sum_{j=1}^{m} \frac{\hat{F}(Y_j)\eta_j}{1-\hat{Q}(Y_j)} - \Delta \\
&= \frac{1}{\sqrt{m}} \int (\hat{F}(Y_j) - \Delta) d\hat{G}(t) \\
&= \frac{1}{\sqrt{m}} \Big[\int_0^x \hat{F}(t) d\hat{G}(t) - \Delta \Big] \\
&= \frac{1}{\sqrt{m}} (\hat{\Delta} - \Delta) \\
&\xrightarrow{\mathcal{D}} N(0, \sigma^2),
\end{aligned}
$$

i.e., $\sum_{j=1}^{m} m_{n_j} \xrightarrow{\mathcal{D}} U$, where $U \sim N(0, \sigma^2)$. Thus, (A1) holds. $\square$

Next let us prove (A2).

Lemma C.2.

$$\hat{\sigma}_1^2 = \frac{1}{m} \sum_{j=1}^{m} \hat{Z}_j^2 + o(p),$$

where $\hat{Z}_j = \dfrac{(\hat{F}(Y_j) - \Delta)\eta_j}{1 - \hat{Q}(Y_j)}$, *and* $\hat{\sigma}_2^2 = \dfrac{(\hat{F}(Y_j) - \hat{\Delta})\eta_j}{1 - \hat{Q}(Y_j)}$.

Proof of C.2.

$$
\begin{aligned}
\frac{1}{m} \sum_{j=1}^{m} \hat{Z}_j^2 &= \frac{1}{m} \sum_{j=1}^{m} \frac{(\hat{F}(Y_j) - \Delta)^2 \eta_j^2}{[1 - \hat{Q}(Y_j)]^2} \\
&= \frac{1}{m} \sum_{j=1}^{m} \frac{\hat{F}(Y_j)^2 - 2\Delta \hat{F}(Y_j) + \Delta^2}{[1 - \hat{Q}(Y_j)]} [\hat{G}(Y_j) - \hat{G}(Y_j-)] \\
&= \int_0^{\infty} \frac{\hat{F}(t)^2 - 2\Delta\hat{F}(t) + \Delta^2}{[1 - \hat{Q}(t)]} d\hat{G}(t) \\
&\xrightarrow{P} \int_0^{\infty} \frac{F(t)^2 - 2\Delta F(t) + \Delta^2}{1 - Q(t)} dG(t) = \sigma_1^2.
\end{aligned}
$$

Let

$$
\begin{aligned}
\Big|\sum_{j=1}^{m}(\frac{1}{\sqrt{m}}\hat{Z}_j)^2-\hat{\sigma}_1^2\Big| =&\Big|\frac{1}{m}\sum_{j=1}^{m}\frac{(\hat{F}(Y_j)-\Delta)^2-(\hat{F}(Y_j)-\hat{\Delta})^2}{1-\hat{Q}(Y_j)^2}\eta_j^2\Big| \\
=&\Big|\frac{1}{m}\sum_{j=1}^{m}\frac{(\Delta-\hat{\Delta})(2\hat{F}(Y_j)-\hat{\Delta}-\Delta)}{1-\hat{Q}(Y_j)^2}\eta_j^2\Big| \\
\leq&|\Delta-\hat{\Delta}|\Big(\frac{1}{m}\sum_{j=1}^{m}\frac{\eta_j^2}{1-\hat{Q}(Y_j)^2}\Big)\sup_{Y_j}\Big|2\hat{F}(Y_j)-\hat{\Delta}-\Delta\Big|.
\end{aligned}
$$

For any Y_j, $|2\hat{F}(Y_j)-\hat{\Delta}-\Delta|\leq 4$, one has

$$
\begin{aligned}
\Big|\sum_{j=1}^{m}(\frac{1}{\sqrt{m}}\hat{Z}_j)^2-\hat{\sigma}_1^2\Big| \leq&4|\Delta-\hat{\Delta}|\frac{1}{m}\sum_{j=1}^{m}\Big[\frac{\eta_j}{1-\hat{Q}(Y_j)}\Big]^2 \\
=&4|\Delta-\hat{\Delta}|\Big\{E(\frac{\eta_j}{1-\hat{Q}(Y_j)})^2+o_p(1)\Big\}.
\end{aligned}
$$

Since $\hat{\Delta}_n\to\Delta$, and $E\dfrac{\eta_j}{1-\hat{Q}(Y_j)}<\infty$, we have

$$
\hat{\sigma}_1^2=\frac{1}{m}\sum_{j=1}^{m}\hat{Z}_j^2+o(p).
$$

Let $a_n=1$, then

$$
a_n\sum_{j=1}^{m}m_{n_j}^2=\sum_{j=1}^{m}\Big(\frac{1}{\sqrt{m}}\hat{Z}_j\Big)^2=\frac{1}{m}\sum_{j=1}^{m}\hat{Z}_j^2\xrightarrow{pr}\hat{\sigma}_1^2.
$$

The condition (A2) holds. □

For condition (A3), we have

$$
\max_i\Big\|\frac{1}{\sqrt{m}}\hat{Z}_j\Big\|=\frac{1}{\sqrt{m}}\max_i\|\hat{Z}_j\|=o(p),
$$

i.e., $\max_i\Big\|\dfrac{1}{\sqrt{m}}\hat{Z}_j\Big\|\xrightarrow{pr}0$. Denote $N=N(0,1)$, then $U=\sigma_1 N\sim N(0,\sigma_1^2)$. Denote $V_2=\hat{\sigma}_1^2$, then

$$
\frac{1}{\hat{\sigma}_1^2}U^2=\frac{1}{\hat{\sigma}_1^2}\sigma^2N^2=\frac{\sigma^2}{\hat{\sigma}_1^2}N^2\sim\frac{\sigma^2}{\hat{\sigma}_1^2}\chi_1^2.
$$

Using Theorem C.1, we have

$$\hat{l}(\Delta_0) \xrightarrow{\mathcal{D}} \gamma(\Delta_0)\chi_1^2,$$

where $\gamma = \dfrac{\sigma^2}{\sigma_1^2}$. That is, Theorem 4.2 is proved.

Appendix D

PROOFS OF CHAPTER 5

For the notations of X, Y, ξ, and η, please refer to Chapter 5. Denote

$$\hat{w}_{1i} = \frac{\hat{F}_1(X_{2i}) - A_1}{1 - \hat{Q}_1(X_{2i})}\xi_{2i};\ \hat{w}_{2j} = \frac{\hat{F}_2(Y_{2j}) - A_1 + D]}{1 - \hat{Q}_2(Y_{2j})}\eta_{2j}.$$

Define the difference of two AUC's as $D = A_1 - A_2$, and

$$\hat{\alpha}_{1i} = \frac{\partial \hat{w}_{1i}}{\partial \theta} = -\frac{\xi_{2i}}{1 - \hat{Q}_1(X_{2i})},\ \hat{\alpha}_{2j} = \frac{\partial \hat{w}_{2j}}{\partial \theta} = -\frac{\eta_{2j}}{1 - \hat{Q}_2(Y_{2j})}.$$

We will prove the assumptions in Valeinis (2007). Similar to Appendix B, we prove (A1)-(A4) for Assumption 4.

Proof of (A1). 1. ($E\hat{w}_1^2 > 0$.)

$$E\hat{w}_1^2 = E\Big\{\Big[\frac{\hat{F}_1(X_{2i}) - A_1}{1 - \hat{Q}_1(X_{2i})}\xi_{2i}\Big]^2\Big\} = E\Big\{\frac{[\hat{F}_1(X_{2i}) - A_1]^2}{[1 - \hat{Q}_1(X_2)]^2}|X_1^0 \leq U_1\Big\}.$$

Since $P(X_1 \leq U_1) > 0$, and $\Big(\frac{\hat{F}_1(X_{2i}) - A_1}{1 - \hat{Q}_1(X_2)}\Big)^2 > 0$, we have

$$E\Big\{\Big(\frac{\hat{F}_1(X_{2i}) - A_1}{1 - \hat{Q}_1(X_2)}\Big)^2|X_2\Big\} > 0.$$

Thus $E\hat{w}_1^2 > 0$.

2. ($\hat{\alpha}_1$ is continuous in a neighbourhood of A_1.)

$\hat{\alpha}_1$ is a constant for A_1, that is, $\hat{\alpha}_1$ is continuous in a neighbourhood of A_1.

4. ($E\hat{\alpha}_1$ is nonzero.)

$$E\hat{\alpha}_1 = E\Big(\frac{-1}{1-\hat{Q}_1(X_2)}\xi_2\Big) = E\Big(\frac{-1}{1-\hat{Q}_1(X_2)}|X_1 \le U_1\Big).$$

Since $\dfrac{-1}{1-\hat{Q}_1(X_2)} < 0$, $E(\dfrac{-1}{1-\hat{Q}_1(X_2)}|X_1) < 0$. $P(X_2 \le U_2) > 0$, then $E\hat{\alpha}_1 < 0$ and $E\hat{\alpha}_1$ is nonzero. □

Proof of (A2) and (A3). Let $\bar{H} = \{F_1, Q_1\}$, $P(\hat{F}_1 = F_1, \hat{Q}_1 = Q_1) \to 1$, by Wang (1987). For $\eta \in (1/3, 1/2)$, $c < \infty$,

$$\begin{aligned}
&\sup_{A_1+\Delta, |\Delta|\le cn^{-\eta}} \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1-\Delta}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - E\frac{F_1(X_{2i})-A_1-\Delta}{1-\hat{Q}_1(X_2)}\xi_2\Big| \\
=&\sup_{A_1+\Delta, |\Delta|\le cn^{-\eta}} \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1-\Delta}{1-\hat{Q}_1(X_{2i})}\xi_{2i} + \Delta\Big| \\
=&\sup_{|\Delta|\le cn^{-\eta}} \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i} + \frac{1}{n}\sum_{i=1}^{n}\Delta(1-\frac{\xi_{2i}}{1-Q_1(X_{2i})})\Big| \\
\le&\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big| + \sup_{|\Delta|\le cn^{-\eta}}|\Delta|\Big|\frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_{2i}}{1-Q_1(X_{2i})})\Big| \\
\stackrel{\Delta}{=}&D_1 + D_2.
\end{aligned}$$

By Marcinkiewicz-Zygmund strong law of large number,

$$\begin{aligned}
\frac{1}{n^{p_1}}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - E\Big(\frac{1}{n^{p_1}}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big) =&\frac{1}{n^{p_1}}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - 0 \\
\to&0 \ a.s.
\end{aligned}$$

for $\frac{1}{2} < p_1 \le 1$.

$$D_1 = n^{1-p_1}\Big[\frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big] = o(n^{p_1-1}) \ a.s.$$

for $0 \le 1-p_1 < \frac{1}{2}$. On the other hand, we have

$$D_2 = \sup_{|\Delta|\le cn^{-\eta}} |\Delta| \Big| \frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_{2i}}{1-Q_1(X_{2i})})\Big|$$
$$\le cn^{-\eta}\Big|\frac{1}{n}\sum_{i=1}^{n}(1-\frac{\xi_{2i}}{1-Q_1(X_{2i})})\Big|.$$

By the SLLN, $\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-Q_1(X_{2i})}-1\Big| = o(1)$ a.s. and $cn^{-\eta} = O(n^{-\eta})$. Thus $D_2 \le O(n^{-\eta})o(1) = o(n^{-\eta})$. That is, $D_2 = O(\beta_0)$, where $\beta_0 = o(n^{-\eta})$. Set $\eta_0 = \eta \in (1/3, 1/2) \subset (0, 1/2)$, then $D_1 = o(n^{\eta})$.

$$\sup_{|\Delta|\le cn^{-\eta}}\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - E\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_2\Big| = O(\beta_1),$$

where $\beta_1 = o(n^{-\eta})$. Thus we proved that

$$w_1 = \frac{1}{n}\sum_{i=1}^{n}\frac{F_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}$$

has the strong Gilvenko-Cantelli property. For $p \in (0,1)$, w_1 is bounded. By Dudley (1998), both w_1^2 and w_1^3 have the strong Gilvenko-Cantelli property.

Next, we prove that g is bounded, $\|g\|_\infty < \infty$, i.e., $\sup_x |g(x)| = \|g\|_\infty < \infty$. Denote

$$g_1 = g_1^* \cdot I_{[-M,M]}(\cdot),\ g_1^* = x^2,\ \ g_2 = g_2^* \cdot I_{[-M,M]}(\cdot),\ g_2^* = x^3.$$

Since w_1^3 is assumed to be bounded, w_1 is bounded as well. Suppose $w_1^3(X,\theta,\hat{h})$ is bounded by an integrable function $G_1(x)$ in the neighbourhood of θ_0, w_1 is bounded by $G_1^{\frac{1}{3}}(x)$. By Stute and Wang (1993), we have

$$\sup_x |\hat{F}_1(x) - F_1(x)| \to 0\ a.s.;\ \sup_x |\hat{Q}_1(x) - Q_1(x)| \to 0\ a.s.$$

Then $\forall h$, as $\hat{h} \to h$, we have $w_1^3(X,\theta,\hat{h}) \le G_1(x)$. Thus $w_1^3(X,\theta,h) \le G_1(x)$. If $G(x)$ is integrable, then $G(x)$ is bounded $a.s.$ Thus $w_1^3(X,\theta,h)$ is bounded $a.s.$ $\square$

Since $\alpha_1 = -\dfrac{\xi_2}{1-\hat{Q}_1(X_2)}$ does not vary as $R_1(p)$ or Δ changes,

$$\sup_{|\Delta|\le cn^{-\eta}} \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_{2i}}{1-Q_1(X_{2i})} - E\frac{\xi_2}{1-\hat{Q}_1(X_2)}\Big| = -\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-Q_1(X_{2i})} + 1.$$

By Marcinkiewicz-Zygmund strong law of large number,

$$1-\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-Q_1(X_{2i})} = O(\beta_1'),$$

where $\beta_1' = o(n^{-\eta_2})$, $\eta_2 \in (1/3, 1/2)$. Thus, w_1, w_1^2, w_1^3, and α_1 have the strong Gilivenko-Cantelli property. Now we rewrite $\hat{E}w_1$ as

$$\begin{aligned} E\hat{w}_1 =& \Big(E\hat{w}_1 - \frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i}\Big) + \Big(\frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big) + \Big(\frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big) \\ \stackrel{\Delta}{=}& E_1 + E_2 + E_3. \end{aligned}$$

By Marcinkiewiz-Zygmund strong law of large number,

$$|E_1| = \Big|E\hat{w}_1 - \frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i})\Big| = O(\beta_2^{(1)}),$$

where $\beta_2^{(1)} = o(n^{-\eta_3^{(1)}})$, $\eta_3^{(1)} \in (1/3, 1/2)$.

$$|E_3| = \Big|\frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big| = \Big|0 - \frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big| = \Big|Ew_1 - \frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big| = O(\beta_2^{(2)}),$$

where $\beta_2^{(2)} = o(n^{-\eta_3^{(2)}})$, $\eta_3^{(2)} \in (1/3, 1/2)$. Next, we define

$$\hat{w}_{1i} = \frac{\hat{F}_1(X_{2i}) - A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i},\ \tilde{w}_{1i} = \frac{\hat{F}_1(X_{2i}) - A_1}{1-Q_1(X_{2i})}\xi_{2i},\ w_{1i} = \frac{F_1(X_{2i}) - A_1}{1-Q_1(X_{2i})}\xi_{2i}.$$

Then E_2 can be rewritten as

$$\begin{aligned} E_2 =& \frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n}w_{1i} \\ =& \frac{1}{n}\sum_{i=1}^{n}\hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{1i} + \frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n}w_{1i}, \end{aligned}$$

where

$$\begin{aligned} \Big|\frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big| =& \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\hat{F}_1(X_{2i}) - F_1(X_{2i})}{1 - Q_1(X_{2i})}\xi_{2i}\Big| \\ \leq& \frac{1}{n}\sum_{i=1}^{n}\frac{|\hat{F}_1(X_{2i}) - F_1(X_{2i})|}{1 - Q_1(X_{2i})}\xi_{2i} \\ \leq& \max_i |\hat{F}_1(X_{2i}) - F_1(X_{2i})|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1 - Q_1(X_{2i})}, \end{aligned}$$

by Holder's inequality. By Smirnov's result in Shorack and Wellner (1986), and

$$\sqrt[3]{n} < \sqrt{\frac{2n}{ln(ln(n))}} < \sqrt{n},$$

when n is large enough, we have

$$\lim_{n\to\infty}\sup\sqrt{\frac{2n}{ln(ln(n))}}\sup_x |\hat{F}_1(x) - F_1(x)| = 1 \ a.s.$$

Thus $\sup_x |\hat{F}_1(x) - F_1(x)| = O_p(\beta_2^{(3)})$, where $\beta_2^{(3)} = o(n^{-\eta})$, $\eta \in (1/3, 1/2)$. By the SLLN,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1 - Q_1(X_{2i})} \to E\frac{\xi_2}{1 - Q_1(X_2)} = 1 \ a.s.$$

Thus

$$\Big|\frac{1}{n}\sum_{i=1}^{n}\tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^{n}w_{1i}\Big| = O(\beta_2^{(3)}) \ a.s.$$

Next consider $\frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i}$. We know that $\frac{1}{n}\sum_{i=1}^n w_{1i} = O(\beta_2^{(2)})$,

$$\frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} = O(\beta_2^{(5)}),\ a.s.,$$

where $\beta_2^{(5)} = o(n^{-\eta_3^{(5)}})$, $\eta_3^{(2)} \in (1/3, 1/2)$. Thus,

$$\begin{aligned}\Big|\frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i}\Big| =& \frac{1}{n}\sum_{i=1}^n \Big\{\frac{\hat{F}_1(X_{2i}) - A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i} - \frac{\hat{F}_1(X_{2i}) - A_1}{1-Q_1(X_{2i})}\xi_{2i}\Big\}\\ &\le \sup_{s\le X_{(n)}}\Big|\frac{\hat{Q}_1(s)-Q_1(s)}{1-\hat{Q}_1(s)}\Big|\Big|\frac{1}{n}\sum_{i=1}^n \frac{\hat{F}_1(X_{2i}) - A_1}{1-Q_1(X_{2i})}\xi_{2i}\Big|,\end{aligned}$$

where

$$\sup_{s\le X_{(n)}}\Big|\frac{\hat{Q}_1(s)-Q_1(s)}{1-\hat{Q}_1(s)}\Big| = O(1),$$

$$\Big|\frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i}\Big| = \Big|\frac{1}{n}\sum_{i=1}^n \frac{\hat{F}_1(X_{2i}) - A_1}{1-Q_1(X_{2i})}\xi_{2i}\Big| = O(\beta_2^{(5)}).$$

$$\frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} = O(\beta_2^{(5)}).$$

Therefore,

$$\begin{aligned}E_2 =& \frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n w_{1i}\\ =& \frac{1}{n}\sum_{i=1}^n \hat{w}_{1i} - \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} + \frac{1}{n}\sum_{i=1}^n \tilde{w}_{1i} - \frac{1}{n}\sum_{i=1}^n w_{1i}\\ =& O(\beta_2^{(6)}),\end{aligned}$$

where $\beta_2^{(6)} = o(n^{-\eta_3^{(6)}})$, $\eta_3^{(6)} \in (1/3, 1/2)$. Moreover, $E\hat{w}_1 = E_1 + E_2 + E_3 = O(\beta_2)$, where $\beta_2 = o(n^{-\eta_3})$, $\eta_3 \in (1/3, 1/2)$. Assumption 4 in Valeinis (2007) is proved.

Similar to Appendix B, we consider Assumption 5. (B1) and (B2) can be proved similar to the one sample case, that is, the AUC's with right censored data.

Proof of (B3). $\hat{Q}_1(X_{2i})$ is the Kaplan-Meier (K-M) estimator of Q, thus $\hat{Q}_1(X_2) \xrightarrow{p} Q_1(X_2)$

point-wise, and

$$\frac{\xi_2}{1-\hat{Q}_1(X_2)} \xrightarrow{\mathcal{D}} \frac{\xi_2}{1-Q_1(X_2)}.$$

By the law of large number (LLN),

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_2}{1-\hat{Q}_1(X_2)} = -E\frac{\xi_2}{1-Q_1(X_2)} + o_p(1).$$

Assume $\hat{\alpha} = \dfrac{\xi_2}{1-\hat{Q}_1(X_2)}$ is bounded by an absolute integrable function, then

$$E\frac{\xi_2}{1-\hat{Q}_1(X_2)} = E\frac{\xi_2}{1-Q_1(X_2)} = 1,$$

by the Lebesgue's Dominated Convergence Theorem. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_{2i}}{1-\hat{Q}_1(X_{2i})} = -1 + o_p(1),$$

that is,

$$\frac{1}{n}\sum_{i=1}^{n}\frac{-\xi_{2i}}{1-\hat{Q}_1(X_{2i})} \xrightarrow{p} -1.$$

□

Proof of (B4). The bounded conditions in (B4) can be directly proved by the assumptions as those in (A1). □

Proof of (B5). (1) For $R_1^*(p) \in (R_1(p) - c_0 n^{-\eta}, R_1(p) + c_0 n^{-\eta})$, $c_0 > 0$,

$$\begin{aligned}
\Big|\frac{\partial}{\partial A_1}\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\hat{F}_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\Big]^2\xi_{2i}\Big| =& 2\Big|\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\hat{F}_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big]\Big| \\
\le& 2\max_i\Big|\frac{\hat{F}_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big|\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big| \\
\triangleq& S_1 + S_2,
\end{aligned}$$

where

$$S_1 = 2\max_i \Big|\frac{\hat{F}_1(X_{2i}) - A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big|,\ S_2 = \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big|.$$

By the conclusion in (B3),

$$S_2 = \Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big| \xrightarrow{p} 1.$$

$$S_1 \le 2\max_i |\hat{F}_1(X_{2i}) - A_1| \max_i \Big|\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big|,$$

where $\max_i \Big|\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big|$ is bounded by a constant $c_{00} \in R$ as given in the assumptions in (B4) and (A1).

$$\begin{aligned} 2\max_i |\hat{F}_1(X_{2i}) - A_1| =& 2\max_i |\hat{F}_1(X_{2i}) - A_1 - cn^{-\eta}| \\ \le& 2\{\max_i |\hat{F}_1(X_{2i})| + \max|A_1| + \max|cn^{-\eta}|\} \\ \le& 2(1+1+c_0 n^{-\eta}) \\ =& 4 + 2c_0 n^{-\eta}, \end{aligned}$$

where $c \in (-c_0, c_0)$, $\eta \in (1/3, 1/2)$. Thus, $S_1 \le c_{00} = 4 + 2c_0 n^{-\eta}$, i.e., $S_1 = O_p(1)$. Therefore,

$$\Big|\frac{\partial}{\partial A_1}\frac{1}{n}\sum_{i=1}^{n}\Big[\frac{\hat{F}_1(X_{2i}) - A_1}{1-\hat{Q}_1(X_{2i})}\Big]^2 \xi_{2i}\Big| = O_p(1).$$

(2) By the conclusions in (B3), we have $\Big|\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big| = O_p(1)$.

(3)

$$\begin{aligned}\Big|\frac{1}{n}\sum_{i=1}^{n}\hat{w}_1^3\Big| =&\Big|\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\hat{F}_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big)^3\Big|\\ \le& \max_i|\hat{F}_1(X_{2i})-A_1|^3\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big)^3\\ \le&(\max_i\hat{F}_1(X_{2i})+\max_i A_1+c_0n^{\eta})^3\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big)^3\\ \le&(2+c_0n^{-\eta})^3\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big)^3\\ \le&(2+c_0n^{-\eta})^3\Big[\max_i\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\Big]^2\frac{1}{n}\sum_{i=1}^{n}\frac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})},\end{aligned}$$

where $c\in(-c_0,c_0)$, $\eta\in(1/3,1/2)$. Then $\dfrac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}$ is bounded and $\dfrac{1}{n}\sum_{i=1}^{n}\dfrac{\xi_{2i}}{1-\hat{Q}_1(X_{2i})}\xrightarrow{p}1$ from (B3), then

$$\frac{1}{n}\hat{w}_1^3=\frac{1}{n}\sum_{i=1}^{n}\Big(\frac{\hat{F}_1(X_{2i})-A_1}{1-\hat{Q}_1(X_{2i})}\xi_{2i}\Big)^3=O_p(1).$$

$\square$

The proofs of the rest of the assumptions can be found in Qin and Zhao (2000). Thus Theorem 5.1 is proved.

Appendix E

PROOFS OF CHAPTER 6

The variance $Var(U_n)$ can be estimated by a consistent estimator $\hat{\sigma}^2$ as in Sen (1960) and Arvesen (1969),

$$\hat{\sigma}^2 = \frac{1}{n_1(n_1-1)}\sum_{i=1}^{n_1}(V_{i,0,0}-\bar{V}_{\cdot,0,0})^2+\frac{1}{n_2(n_2-1)}\sum_{j=1}^{n_2}(V_{0,j,0}-\bar{V}_{0,\cdot,0})^2+\frac{1}{n_3(n_3-1)}\sum_{k=1}^{n_3}(V_{0,0,k}-\bar{V}_{0,0,\cdot})^2.$$

Lemma E.1. *Assume that*

(a) The U-statistic $U_n \overset{a.s.}{\to} \theta$ *as* $min(n_1, n_2, n_3) \to \infty$*;*

(b) Assume that $\sigma^2_{1,0,0} > 0$, $\sigma^2_{0,1,0} > 0$, $\sigma^2_{0,0,1} > 0$ *and denote* $S^2_{n_1,n_2,n_3} = \sigma^2_{1,0,0}/n_1 + \sigma^2_{0,1,0}/n_2 + \sigma^2_{0,0,1}/n_3$*. Then*

$$\frac{U_n - \theta}{S_{n_1,n_2,n_3}} \overset{d}{\to} N(0,1), \quad as \quad min(n_1, n_2, n_3) \to \infty, \tag{E.1}$$

and

$$\hat{\sigma}^2 - S^2_{n_1,n_2,n_3} = o_p((min(n_1, n_2, n_3))^{-1}). \tag{E.2}$$

For the proof of part (a) and equations (E.1) and (E.2) in part (b), we may refer to Arvesen (1969) and Kowalski and Tu (2007).

Lemma E.2. *Let* $S_n = n^{-1}\sum_{l=1}^{n}(\hat{V}_l - E\hat{V}_l)^2$*. We assume the same conditions as (a) and (b) in Theorem 6.1. Then as* $n_1 \to \infty$*,*

$$S_n = nS^2_{n_1,n_2,n_3} + o_p(1), \ a.s.$$

Proof of Lemma E.2. For $1 \le l \le n_1$, we have

$$\hat{V}_l - E\hat{V}_l = \frac{n(n-1)}{(n-3)n_1}(V_{l,0,0} - U_n) + \frac{n(n-2n_1-1)}{(n-3)n_1}(U_n - \theta),$$

and

$$
\begin{aligned}
&\frac{1}{n_1}\sum_{l=1}^{n_1}(V_{l,0,0}-U_n)(U_n-\theta)\\
=&(U_n-\theta)\{\frac{1}{n_1n_2n_3}\sum_{i=1}^{n1}\sum_{j=1}^{n2}\sum_{k=1}^{n3}[I(X_{1i}<Y_{1j}<Z_{1k})-I(X_{2i}<Y_{2j}<Z_{2k})]-U_n\}\\
=&0.
\end{aligned}
$$

Thus,

$$
\sum_{l=1}^{n_1}(\hat{V}_l-E\hat{V}_l)^2=\Big[\frac{n(n-1)}{(n-3)n_1}\Big]^2\sum_{l=1}^{n_1}(V_{l,0,0}-U_n)^2+\Big[\frac{n(n-2n_1-1)}{(n-3)n_1}\Big]^2 n_1(U_n-\theta)^2.
$$

Similarly, for $n_1 < l \le n_1+n_2$,

$$
\sum_{l=n_1+1}^{n_1+n_2}(\hat{V}_l-E\hat{V}_l)^2=\Big[\frac{n(n-1)}{(n-3)n_2}\Big]^2\sum_{l=n_1}^{n_1+n_2}(V_{0,l,0}-U_n)^2+\Big[\frac{n(n-2n_2-1)}{(n-3)n_2}\Big]^2 n_2(U_n-\theta)^2.
$$

And for $n_1+n_2 < l \le n$,

$$
\sum_{l=n_1+n_2+1}^{n}(\hat{V}_l-E\hat{V}_l)^2=\Big[\frac{n(n-1)}{(n-3)n_3}\Big]^2\sum_{l=n_1+n_2}^{n}(V_{0,0,l}-U_n)^2+\Big[\frac{n(n-2n_3-1)}{(n-3)n_3}\Big]^2 n_3(U_n-\theta)^2.
$$

Therefore,

$$
\begin{aligned}
S_n=&\frac{1}{n}\Big[\frac{n(n-1)}{(n-3)}\Big]^2\Big[\frac{1}{n_1^2}\sum_{l=1}^{n_1}(V_{l,0,0}-\bar{V}_{\cdot,0,0})^2+\frac{1}{n_2^2}\sum_{l=n_1+1}^{n_1+n_2}(V_{0,l,0}-\bar{V}_{0,\cdot,0})^2+\frac{1}{n_3^2}\sum_{l=n_1+n_2+1}^{n}(V_{0,0,l}-\bar{V}_{0,0,\cdot})^2\Big]\\
&+\frac{1}{n}\Big[\frac{n}{(n-3)}\Big]^2\Big[\frac{(n-2n_1-1)^2}{n_1}+\frac{(n-2n_2-1)^2}{n_2}+\frac{(n-2n_3-1)^2}{n_3}\Big](U_n-\theta)^2.
\end{aligned}
\tag{E.3}
$$

From the LLN of U-statistics, we have $U_n-\theta=O_p(n_1^{-1/2})$. Hence, the second term in equation (E.3) is equal to

$$
\frac{n}{(n-3)^2}\Big[\frac{(n-2n_1-1)^2}{n_1}+\frac{(n-2n_2-1)^2}{n_2}+\frac{(n-2n_3-1)^2}{n_3}\Big](U_n-\theta)^2=O_p(n^{-1}).
$$

Moreover, the 1st term of equation (E.3) is equal to

$$n(\frac{n-1}{n-3})^2\Big[\frac{1}{n_1^2}\sum_{l=1}^{n_1}(V_{l,0,0}-\bar{V}_{\cdot,0,0})^2+\frac{1}{n_2^2}\sum_{l=n_1+1}^{n_1+n_2}(V_{0,l,0}-\bar{V}_{0,\cdot,0})^2+\frac{1}{n_3^2}\sum_{l=n_1+n_2+1}^{n}(V_{0,0,l}-\bar{V}_{0,0,\cdot})^2\Big]$$

$$=n\hat{\sigma}^2+o_p(1).$$

Note that $\hat{\sigma}^2 - S^2_{n_1,n_2,n_3} = o_p((min(n_1,n_2,n_3))^{-1})$, we prove Lemma E.2. □

Lemma E.3. *Let* $Q_n = \max_{1\le l\le n}|\hat{V}_l-\theta|$. *Under the assumptions as in Lemma E.2, we have* $Q_n = o_p(n^{1/2})$ *a.s. and* $n^{-1}\sum_{l=1}^{n}|\hat{V}_l-\theta|^3 = o_p(n^{1/2})$, *a.s.*

Proof of Lemma E.3. For $1\le l\le n_1$, we have

$$|\hat{V}_l - E\hat{V}_l| \le \Big|\frac{n}{n_1}\frac{n-1}{n-3}V_{l,0,0}\Big| + \Big|\frac{n}{n_1}\frac{n-1}{n-3}U_n\Big| + \Big|\frac{n(n-2n_1-1)}{(n-3)n_1}(U_n-\theta)\Big|.$$

Note that $|V_{l,0,0}|\le \tilde{H}_n$, and $|U_n|\le\tilde{H}_n$, where

$$\tilde{H}_n = \max_{1\le i\le n_1<l\le n_1+n_2<k\le n}|h(X_i,Y_j,Z_k)|.$$

Therefore, $|\hat{V}_l - E\hat{V}_l| \le c^*\tilde{H}_n + c^*\tilde{H}_n + c^*|U_n-\theta|$, where c^* is a constant. Similar to Wang (2010), $\tilde{H}_n = o_p(n^{1/2})$ a.s. and $U_n-\theta = O_p(n^{-1/2})$. Hence, $|\hat{V}_l - E\hat{V}_l| = o_p(n^{1/2})$ for $1\le l\le n_1$. Similarly, for $n_1<l\le n_1+n_2$, $|\hat{V}_l-E\hat{V}_l|\le 2c^*\tilde{H}_n + c^*|U_n-\theta|$.Thus, $|\hat{V}_l-E\hat{V}_l| = o_p(n^{1/2})$ for $n_1<l\le n_1+n_2$. And, for $n_1+n_2<l\le n$, $|\hat{V}_l-E\hat{V}_l|\le 2c^*\tilde{H}_n + c^*(U_n-\theta)|$. Therefore, $|\hat{V}_l - E\hat{V}_l| = o_p(n^{1/2})$ for $n_1+n_2<l\le n$. Denote $Q_n = \max_{1\le i\le n}|\hat{V}_i-\theta|$, then $Q_n = o_p(n^{1/2})$ a.s. Therefore,

$$n^{-1}\sum_{l=1}^{n}|\hat{V}_l - E\hat{V}_l|^3 = o_p(n^{1/2})(nS^2_{n_1,n_2,n_3}+o_p(1)) = o_p(n^{1/2}).$$

□

Proof of Theorem 6.1. Recall $U_n = \frac{1}{n}\sum_{l=1}^{n}\hat{V}_l$ and $\theta = \frac{1}{n}\sum_{l=1}^{n}E\hat{V}_l$. Then $|U_n-\theta| =$

$\Big|\frac{1}{n}\gamma\sum_{l=1}^{n}\frac{(\hat{V}_l-E\hat{V}_l)^2}{1+\gamma(\hat{V}_l-E\hat{V}_l)}\Big|$. Thus,

$$|U_n-\theta|\geq|\gamma|\frac{1}{1+|\gamma|\max|\hat{V}_l-E\hat{V}_l|}\frac{1}{n}\sum_{l=1}^{n}(\hat{V}_l-E\hat{V}_l)^2\geq|\gamma|\frac{S_n}{1+|\gamma|Q_n}.$$

Then $|\gamma|=O_p(n^{-1/2})$. By Taylor's expansion,

$$-2\log R(\theta)=2\sum_{l=1}^{n}\{\gamma(\hat{V}_l-E\hat{V}_l)-\frac{1}{2}[\gamma(\hat{V}_l-E\hat{V}_l)]^2\}+o_p(1). \tag{E.4}$$

Let $F_0=\frac{1}{n}\sum_{l=1}^{n}\frac{\gamma^2(\hat{V}_l-E\hat{V}_l)^3}{1+\gamma(\hat{V}_l-E\hat{V}_l)}$. Then by equation (6.2), we have $\gamma=\frac{U_n-\theta}{S_n}+\frac{F_0}{S_n}$. In equation (E.4), we replace γ with the above terms,

$$2n\gamma(U_n-\theta)-nS_n\gamma^2=n\frac{(U_n-\theta)^2}{S_n}-\frac{nF_0^2}{S_n}.$$

Combining $\frac{1}{n}\sum_{l=1}^{n}|\hat{V}_l-E\hat{V}_l|^3=o_p(n^{1/2})$ and $\gamma=O_p(n^{-1/2})$, $F_0=o_p(n^{-1/2})$. Thus,

$$-2\log R(\theta)=n\frac{(U_n-\theta)^2}{S_n}+o_p(1).$$

By Lemma E.1 and Lemma E.2, one has that $\frac{n(U_n-\theta)^2}{S_n}\xrightarrow{d}\chi_1^2$. By the Slutsky's theorem, we finish the proof of Theorem 6.1. $\square$